UNIVERSIDADE DE LISBOA
INSTITUTO de CIÊNCIAS SOCIAIS

Programa de Doutoramento em Migrações

Social and individual factors in witnesses' confrontations of racial bias

Susana Margarida Silva Ferreira Lavado

Orientadores: Prof. Doutor Cicero Roberto Pereira
Professor Doutor John F. Dovidio
Professor Doutor Jorge Manuel Vala Salvador

Tese especialmente elaborada para a obtenção do grau de doutor no ramo de Psicologia, especialidade de Psicologia Social

2016

UNIVERSIDADE DE LISBOA
INSTITUTO de CIÊNCIAS SOCIAIS

Programa de Doutoramento em Migrações

Social and individual factors in witnesses' confrontations of racial bias

Susana Margarida Silva Ferreira Lavado

Orientadores: Prof. Doutor Cicero Roberto Pereira
Professor Doutor John F. Dovidio
Professor Doutor Jorge Manuel Vala Salvador

Júri:
Presidente: Professor Doutor José António Machado da Silva Pais
Vogais: Prof. Doutora Isabel Maria Rocha Pinto
Prof. Doutor Hélder António Vinagreiro Gomes Alves
Professora Doutora Maria Lucinda Cruz dos Santos Fonseca
Prof. Doutor Cícero Roberto Pereira

Tese especialmente elaborada para a obtenção do grau de doutor no ramo de Psicologia, especialidade de Psicologia Social

Instituição Financiadora: Fundação para a Ciência e a Tecnologia Bolsa de Doutoramento SFRH/BD/81811/2011

2016

Acknowledgements

Looking back, I cannot stop thinking what an amazing adventure this was. Sometimes full of disbelief and despair, but mostly full of discoveries and amazement. Through this acknowledgements, you will notice I use the word "inspiration" over and over, but the truth is that is what I feel – very inspired and very grateful for all the amazing people I had the chance to share this journey with.

When you find yourself at the end of many meetings thinking that you have the best advisor in the world, it must be true. Thank you Cicero, for your patience even when I was lost and annoying. For being an optimist even when I thought everything was doomed and I would never graduate. To let me try new things, and sometimes crash and fall and sometimes succeed. And for being an inspiration as a researcher and as a person.

Jack, even though you barely knew anything about me, you were kind enough to give me the amazing opportunity to work with you. I have never learned so much as I did at Yale, in so many levels. I feel very, very grateful for the privilege to have worked with you. Your guidance and your attitude in life will stay with me forever.

Jorge, if it wasn't for your encouragement, I would not have started this adventure. I distinctly remember my first conversations with you about the possibility of doing a PhD. Thank you for believing in me since day zero, when I barely knew what it meant to be a researcher and for always challenging me to be better and aim higher.

I also wish to thank you my Master's adviser, Maria João Alvarez. Your teaching and guidance accompanied me during this PhD, and I will never forget a saying you taught me: "If you think you are too small to make a difference, you never spent a night with a mosquito."

I was constantly amazed by the inspiration given by all the graduate students with whom I shared this path (visiting students in ICS, members of the Yale intergroup lab, EASP summer schoolers). Having met all of you, I know that social psychology is in good hands. A special thanks to Kika, Fernando and Rodrigues – we rocked as the first class of this PhD program!

Thank you to Ana Raquel Torres for receiving me in UFPB – it was an inspiration to meet you and I wish I had the chance to spend more time with you.

People often talk about the importance of being kind to strangers. I was lucky to have a lot of strangers being very kind to me. Thank you to all the Professors who helped to collect data for this dissertation, and all the students who agree to participate

(sometimes in exchange for a chocolate bar, most always out of kindness). Thank you also to João Barreiros, without whom I would not have been able to collect the data for this thesis.

Thank you to Alice for helping me figure out that I did enjoy this 'research thing'. I've learned so much with you! Thank you for your patience and guidance. And a big thank you to Jussara. You never, ever cease to amaze me with your ability to think critically and your practical spirit - I want to have a little bit of those when I grow up!

Because it is often the small gestures than meant a lot, I also wish to thank Goretti Matias, for all her help during this PhD.

Thank you for helping me feel at home far away from home: Anita & João Manuel, Rosemarie & Louise, Chen, Andrea, Suzanne, Nicholas, Seval, Silvia, Johannes, Aric, Leslie, Jimmy, Peggy, Silvana, Iara, Aline and her family.

I met some of the most special people in the world through GASTagus, whom I won't risk to nominate. Thank you for showing me that happiness is in the small things and that the world can be a wonderful place, but also to think outside of the box. I would have never found the end of this thesis without any of those pieces of knowledge. And thank you to all my friends who never let me feel alone in the pains and joys of doing a PhD, who challenged me to push myself and helped me to dream higher; and who never let me forget that there is more world outside this thesis!

They say good things come in small packages. I have a small (that just got bigger – welcome Olivia!), but very, very special family. Thank you for always supporting me and never let me feel I was alone, even when it was hard for me to explain exactly what I was doing.

Thank you Pedro for showing me that if everything in life will eventually pass, I might as well choose to be happy and fearless every single day. You've shared every bad and good moment of this PhD with me – this is also your thesis and I wouldn't have done it without you!

A very special Thank You! to Ana & Luís, who always believed I could become a doctor (although this type of doctor was not what they had in mind). I do not get to say this very often, but I love you guys!

Abstract

Expressions of prejudice against black people and, more specifically, against black immigrants are still prevalent in Portugal and in other western countries (Dickter & Newton, 2013; Santos, Oliveira, Rosário, Kumar & Brigadeiro, 2005). When people witness expressions of prejudice against others, they can decide to confront those expressions, that is, to express dissatisfaction with the prejudiced behavior to the person responsible for that behavior (Kaiser & Miller, 2004). Confronting prejudice has important positive consequences, decreasing future expressions of prejudice from the confronted person (Czopp, Monteith & Mark, 2006) and observers' agreement with the biased expressions (Rasinsky & Czopp, 2010). The positive consequences of confrontation occur especially when the confronter is not the target of the prejudiced behavior (Gulker, Mark & Monteith, 2013).

In the current thesis, we present a research project investigating social and individual factors influencing witnesses' confrontations of bias. In seven studies, we approached the topic from three main perspectives: (a) the social norms regarding witnesses' responses to bias; (b) observers' attitudes regarding witnesses' responses to bias; and (c) witnesses' actual confrontational behavior. We showed that social factors emerging from the relationship between the person expressing bias and the witness of the biased remark influence norms and attitudes toward confrontational behaviors and witnesses' actual responses to bias. More specifically, the results of the first empirical chapter suggest that when the person expressing bias has high power over the potential confronter, observers' attitudes toward confrontational behaviors are less favorable; in the second empirical chapter, results suggest that witnesses tend to confront an ingroup member who expresses bias less strongly than an outgroup member who expresses bias; and in the final empirical chapter, we showed that it is more normative to confront a close person than a stranger. However, individual differences in egalitarian values and standards moderated the effects of these social factors in attitudes toward confrontations of bias and confrontational behaviors. People who have strong egalitarian values and standards have positive attitudes toward confrontation and confront expressions of bias even in unfavorable social conditions. With the present thesis, we contributed to the current knowledge about witnesses' confrontations of bias, an interpersonal process that may fundamentally help reducing of public expressions of prejudice.

Key-words: Racial prejudice, confrontation, witnesses, egalitarian values

Resumo

Expressões de preconceito contra pessoas negras e, mais especificamente, contra imigrantes negros ainda são prevalente em Portugal e noutros países ocidentais (Dickter & Newton, 2013; Santos, Oliveira, Rosário, Kumar & Brigadeiro, 2005). Quando uma pessoa testemunha expressões de preconceito contra outros grupos, esta pode decidir confrontar essas expressões, ou seja, exprimir desagrado com o comportamento preconceituoso diretamente à pessoa responsável pelo comportamento (Kaiser & Miller, 2004). A confrontação de expressões de preconceito tem consequências positivas importantes, diminuindo expressões preconceituosas futuras por parte de quem é confrontado (Czopp, Monteith & Mark, 2006). Além disso, observar outros a confrontar expressões preconceituosas diminui o grau em que outras pessoas apoiam essas expressões (Rasinsky & Czopp, 2010). As consequências positivas da confrontação ocorrem especialmente quando o confrontador não é o alvo do comportamento preconceituoso (Gulker, Mark & Monteith, 2013).

Na presente tese, apresentamos um projeto de investigação que teve como objetivo testar fatores sociais e pessoais que influenciam a confrontação de preconceito por parte das testemunhas. Em sete estudos, abordámos o tópico de três perspetivas distintas: (a) as normas sociais que governam as respostas das testemunhas a expressões de preconceito; (b) as atitudes de observadores face às respostas das testemunhas a expressões de preconceito; e (c) o próprio comportamento das testemunhas de preconceito. Demonstrámos que fatores sociais com origem na relação entre a pessoa que exprime preconceito e a testemunha influenciam as normas e atitudes face à confrontação e a própria resposta da testemunha face à expressão de preconceito. Mais especificamente, no primeiro capítulo empírico demonstrámos que quando a pessoa que exprime preconceito tem poder sobre o potencial confrontador, as atitudes dos observadores face à confrontação são menos favoráveis. No segundo capítulo empírico, os resultados mostraram que as testemunhas estão menos dispostas a confrontar um membro do endogrupo que exprime preconceito do que um membro do exogrupo que exprime preconceito. No último capítulo empírico, os resultados sugerem que é mais normativo confrontar uma pessoa próxima do que um desconhecido. No entanto, diferenças individuais na adesão e internalização de valores igualitários moderam os efeitos destas variáveis sociais nas atitudes face à confrontação e no próprio comportamento de confrontação. As pessoas que têm valores e padrões igualitários altos têm atitudes

positivas face à confrontação e confrontam expressões de preconceito mesmo em condições sociais menos favoráveis. Com a presente tese, contribuímos para o conhecimento científico sobre a confrontação de preconceito por parte das testemunhas, um processo interpessoal fundamental na redução de expressões públicas de preconceito.

Palavras-chave: preconceito racial, confrontação, testemunhas, valores igualitários

Resumo Alargado

Atualmente, muitas sociedades têm normas que condenam a discriminação de imigrantes e de membros de grupos racializados. É o caso da sociedade Portuguesa e da sociedade Americana, onde a discriminação com base na origem nacional ou étnica é condenada, formal e informalmente (Crandall, Eshleman, & O'Brien, 2002; Dovidio & Gaertner, 2000; Katz & Hass, 1988; Vala, Lopes & Lima, 2008; Vala & Pereira, 2012). No entanto, expressões de preconceito contra estes grupos continuam a ser prevalentes no dia-a-dia, principalmente contra pessoas negras (que sejam imigrantes ou não), quer em Portugal quer nos Estados-Unidos (Dickter & Newton, 2013; Santos, Oliveira, Rosário, Kumar & Brigadeiro, 2005; Swim, Hyers, Cohen, Fitzgerald, & Bylsma, 2003), com consequências negativas sérias para o bem-estar das pessoas que são alvo dessas expressões (Paradies et al., 2015; Schmitt, Branscombe, Postmes & Garcia, 2014).

Uma das formas de reduzir estas expressões negativas contra pessoas com diferentes origens nacionais e étnicas é confrontando o preconceito. A confrontação pode ser definida como a expressão de desagrado com o comportamento preconceituoso, feita diretamente à pessoa que expressa preconceito (Kaiser & Miller, 2004; Shelton, Richeson, Salvatore, & Hill, 2006). De fato, estudos anteriores têm demonstrado que a confrontação é eficaz a reduzir futuras expressões de preconceito, a motivar a pessoa que é confrontada a compensar o alvo do seu preconceito e a diminuir a concordância de observadores da confrontação com expressões preconceituosas (Czopp, Monteith & Mark, 2006; Mallet & Wagner, 2011; Rasinski & Czopp, 2010). No entanto, a confrontação pode ser especialmente eficaz quando é realizada por testemunhas, isto é, por pessoas que não são, direta ou indiretamente alvo do comportamento preconceituoso. A maior eficácia da confrontação de preconceito por testemunhas deve-se a dois fatores. Por um lado, a confrontação por parte de testemunhas é mais surpreendente, levando a um maior processamento da mensagem; por outro lado, gera menor reações negativas não só por parte da pessoa que é confrontada mas também por parte de pessoas que observam a confrontação de preconceito (Gulker, Mark & Monteith, 2013; Rasinski & Czopp, 2010). Importa, por isso, compreender os fatores que podem motivar ou desencorajar a confrontação de preconceito por parte de testemunhas.

Na presente tese, procurámos integrar fatores sociais (com origem na relação entre a pessoa que expressa preconceito e a testemunha) e fatores individuais (relacionados com a adesão a valores igualitários) para compreender a confrontação de preconceito

racial por parte de testemunhas sob três perspetivas diferentes: (a) as normas sociais prescritivas relativas à confrontação de preconceito; (b) as atitudes dos observadores face às respostas de confrontação ou não confrontação; e (c) o comportamento de confrontação de preconceito.

Os resultados de sete estudos, organizados nos três capítulos empíricos que compõem esta tese, sugerem que características da relação entre a pessoa que expressa preconceito e a pessoa que testemunha essa expressão de preconceito influenciam o comportamento, as atitudes e as normas sociais relativas à confrontação de preconceito. No entanto, a influência destes fatores sociais nas atitudes dos observadores e nas respostas de confrontação ou não confrontação das testemunhas dependem da adesão a valores igualitários. Pessoas (não-alvo de preconceito) que têm fortes valores igualitários apoiam mais a confrontação de preconceito e estão mais dispostas a confrontar preconceito, mesmo em condições sociais menos favoráveis à confrontação.

No primeiro capítulo empírico, avaliámos as atitudes de observadores do grupo maioritário face à reação de uma testemunha que, no decorrer de uma entrevista de trabalho, confronta ou não confronta um comportamento preconceituoso do seu entrevistador. Manipulámos também o grau de poder do entrevistador face à testemunha de preconceito e, consequentemente, os custos da confrontação: confrontar podia ter baixos custos (porque a entrevista era para um emprego pouco desejado pela testemunha) ou altos custos (porque a entrevista era para um emprego muito desejado pela testemunha). As atitudes dos observadores face à confrontação foram geralmente mais positivas do que as atitudes face à não confrontação, replicando resultados anteriores que sugerem que as testemunhas que confrontam são melhor avaliadas que as testemunhas que não confrontam (Dickter, Kittel, & Gyurovski, 2012). A avaliação da confrontação foi especialmente positiva quando a confrontação tinha custos baixos. Quando confrontar implicava custos altos, o comportamento de confrontação foi mais positivamente avaliado do que o comportamento de não confrontação apenas pelos observadores com forte adesão a valores igualitários. Assim, no primeiro capítulo empírico demonstramos que os observadores com maior adesão a fatores igualitários são menos influenciados por fatores sociais relacionados com os custos da confrontação.

No segundo capítulo empírico investigámos a influência da partilha de uma pertença grupal com a pessoa que exprime preconceito no grau em que testemunhas confrontam esse mesmo preconceito. Uma vez que comportamentos negativos realizados

por membros de um exogrupo tendem a ser mais atribuídos a disposições internas do que comportamentos negativos realizados por membros de um endogrupo (Hewstone, 1990; Petigrew, 1979), colocámos a hipótese que as testemunhas de preconceito iriam atribuir mais características tipicamente associadas a pessoas racistas a um membro de um exogrupo que faz um comentário preconceituoso do que a um membro do endogrupo que faz o mesmo comentário. Por sua vez, quanto mais se atribuem características racistas à pessoa que faz um comentário preconceituoso, mais forte deverá ser a confrontação. Os resultados apoiaram as duas hipóteses. No entanto, o efeito da pertença grupal na atribuição de características associadas a pessoas racistas foi moderado pelo grau em que os participantes tinham internalizado padrões igualitários, ou seja, o grau em que os participantes estavam motivados para responder sem preconceito (Plant & Devine, 1998, 2009). Apenas os participantes menos motivados para responder sem preconceito atribuíram menos características preconceituosas e consequentemente confrontaram menos um membro de um endogrupo do que um membro de um exogrupo. Uma vez que as pessoas mais motivadas para responder sem preconceito são altamente responsivas a expressões de preconceito contra outros grupos, estas confrontaram a pessoa que expressou preconceito independentemente da sua pertença grupal.

Por fim, no último capítulo empírico, analisámos a influência da relação entre a pessoa que faz um comentário preconceituoso e a testemunha desse comentário nas normas prescritivas que regem as respostas a expressões de preconceito. Em linha com o demonstrado no primeiro capítulo empírico, os resultados deste último capítulo empírico mostraram que a confrontação de preconceito é um comportamento valorizado socialmente. No entanto, a confrontação de preconceito foi vista como mais normativa quando a pessoa que expressa preconceito é alguém próximo (isto é, um amigo ou um familiar), comparativamente com quando a pessoa que expressa preconceito é um desconhecido. Expressões de preconceito por parte de pessoas próximas constituem uma maior ameaça à imagem de grupos altamente valorizados (grupos de amigos e família) do que expressões de preconceito de desconhecidos. A ameaça à imagem de um grupo altamente valorizado, por sua vez, legitima comportamentos de confrontação por parte das testemunhas de preconceito, tornando mais normativo confrontar.

Os resultados da presente tese contribuem de forma importante para a literatura sobre a confrontação de preconceito racial. A confrontação de preconceito por parte de testemunhas tem sido pouco estudada na literatura, uma vez que estudos anteriores têm-

se focado principalmente nas respostas de alvos de preconceito. Além disso, investigámos fatores que não tinham sido ainda explorados nos estudos sobre confrontação, tais como a partilha de uma pertença grupal e a ameaça à imagem de um grupo altamente valorizado. Por fim, mostrámos que a influência de fatores sociais (com origem na relação entre a pessoa que expressa preconceito e a testemunha desse preconceito) nas atitudes e comportamentos de confrontação depende da adesão a valores igualitários por parte dos observadores e das testemunhas de preconceito. Desta forma, os nossos resultados contribuem diretamente para o conhecimento científico sobre as respostas de testemunhas de preconceito, comportamentos interpessoais que, como referido, são processos fundamentais na redução ou manutenção de expressões de preconceito racial.

Table of contents

Index of tables and figures

Chapter I: Introduction

Chapter II: Evaluations of Witnesses' Responses to Bias: Universalism-Concern and the Costs of Confrontation

Chapter III: Ingroup favoritism and witnesses' confrontations of racial bias

Chapter IV: Making us look bad: Group image threat and the normativity of witnesses' confrontations of bias

Chapter I

Introduction

Imagine that you hear someone who shares your nationality and/or ethnicity making a prejudiced comment against immigrants or people perceived to be from a different ethnicity. You are aware that social norms in the country condemn expressions of prejudice against those groups, and it is likely that you have also internalized those norms to some extent. Would you confront the person, expressing your disagreement to him or her? Or would you ignore the expression of prejudice? And what do you think other people in the same situation should do? Would you support another person's decision to confront? Because the situation is described in a rather vague manner, your answer to the above questions probably is "it depends". It may depend on who the person who made the comment is, that is, on social factors; but also on your own personal beliefs, that is, on individual differences in value-based motivations.

In the present thesis, we investigated social and individual factors that may influence how witnesses' respond to racial bias and how observers' evaluate witnesses' responses to prejudice. We focused on expressions of prejudice against Black people and, more specifically, against Black immigrants, because they are one of the groups that are more frequently targeted by everyday discrimination in Portugal and the US, the countries where this research was conducted (Dickter & Newton, 2013; Santos, Oliveira, Rosário, Kumar & Brigadeiro, 2005). However, even though our studies were focused on Black people and on Black immigrants, they contribute to the understanding of the psychological processes underlying high-status group members' responses to prejudice targeting any low-status groups who are simultaneously protected by egalitarian norms and the targets of prejudice.

We will next summarize research investigating the prevalence of expressions of bias against Black people in Portugal and in the US. We will then review past work regarding confrontations of bias and present the hypotheses of the current thesis, as well as its chapters' outline.

Expressions of prejudice in Portugal and in the US

Both in Portugal and in the United States, egalitarian norms condemn expressions of prejudice toward Black people, such as immigrants or descendants of immigrants from African countries living in Portugal (e.g., Vala, Brito & Lopes, 1999/2015; Vala, Lopes & Lima, 2008; Vala & Pereira, 2012) or African-Americans in the United States (e.g., Crandall, Eshleman, & O'Brien, 2002; Dovidio & Gaertner, 2000; Katz & Hass, 1988;

Plant & Devine, 1998). In addition to being informally condemn, discrimination based on race, skin color, nationality or ethnic origin is formally punished by the Portuguese law since 1999[1] (complementing the 13th chapter of the Portuguese Constitution of 1974 also prohibiting any type of discrimination based on origin, sex, 'race', language religion, political or ideological beliefs, education, economic position, social class or sexual orientation), while several Federal Laws in the US also prohibit discrimination based on a person's national origin, race, color, religion, disability, sex, and familial status[2]. In an index evaluating countries' policies to integrate migrants, including anti-discrimination policies, Portugal is described as having "favorable [anti-discrimination] laws, enforcement mechanisms and equality policies" but it is also suggested that Portugal's "relatively recent laws and policies may be too poorly known or resourced to get potential victims to regularly report discrimination." According to the same index, in the US "[people] enjoy the strongest laws to protect them against discrimination" (MIPEX - Huddleston, Bilgili, Joki, & Vankova, 2015).

However, despite the laws, policies and norms condemning prejudiced behaviors, Black people in both countries still report – in surveys and interviews – facing everyday racism – that is, they still experience being the target of another person's prejudice and discrimination in mundane social interactions (Essed, 1991). As much as 29% of a sample of people from sub-Saharan African countries living in Portugal[3] reported being discriminated against due to their ethnicity or immigrant status at least once in the last year, in domains such as work, housing and public services (European Union Agency for Fundamental Rights, 2009). Similarly, in a survey about immigrants' experiences in Portugal, immigrants from Cape Verde and Guinea-Bissau reported having been the target of harassment (e.g., being insulted or threatened) in the streets more often than immigrants from other national origins, such as Brazil and Ukraine. In fact, 38% of the respondents from both countries reported having been harassed on the street at least once in the last year. Of these, more than one quarter reported having been publicly harassed

[1] Portugal/Law 134/99 (28.08.1999)
[2] The United States Department of Justice – http://www.justice.gov/crt/federal-protections-against-national-origin-discrimination-1#ed
[3] Immigrants from African countries constitute about 25% of the total population of immigrants legally residing in the country. They are mainly from Portuguese-speaking countries such as Cape-Verde, Angola and Guinea-Bissau (SEF, 2014).

five or more times in the last year (Santos, et al., 2005).[4] A more recent survey also found that discrimination often occurs in public places, such as public transports, stores or cafes and restaurants (Mendes & Candeias, 2013). Non-reporting the experiences of discrimination (at the place they occurred or to a competent authority) is still, by far, the most common behavior of immigrants and ethnic minorities in Portugal (European Union Agency for Fundamental Rights, 2009; Peixe et al., 2008). Even so, in 2014 the Portuguese *Committee for equality and against racial discrimination* [Comissão para a Igualdade e Contra a Discriminação Racial] registered 60 complains of racial discrimination, the majority of which were made by people who self-identified as Black (CICDR - ACIME/ACIDI/ACM, 2016).

Similarly, studies conducted in the United States show that racially biased remarks tend to target Black people more often than other racial groups (Dickter & Newton, 2013) and that Black people report having to deal with other people's prejudice frequently. For instance, in a sample of African-American undergraduate students, only 11% reported never having heard negative remarks about Black people on campus (D'Augelli, & Hershberger, 1993). In a recent poll, 45% of Black interviewees reported being the target of racial discrimination at least once a month (CNN/ORC, 2015). African-American undergraduate students reported on average 1.24 events (*SD* = 1.45) in the course of two weeks that they considered probably or definitely caused by prejudice. Of these, about one quarter were verbal expressions of prejudice, and about one third occurred in a public or institutional setting (Swim, Hyers, Cohen, Fitzgerald, & Bylsma, 2003). Likewise, interviews with Black middle-class people suggest that verbal expressions of prejudice constitute 25% of the reported discriminatory events occurring on the street, and 12% of the reported discriminatory events occurring on public places like shops or restaurants (Feagin, 1991; however, like in the above cited study of Swim et al., 2003, these

[4] Brazilians and Ukrainians reported being more discriminated against in other areas, such as in access to credit and being promoted at work. It should be noted that Mendes and Candeias (2013) survey of the immigrant population residing in the Oeiras municipality found that Brazilians reported being the most discriminated against: 67% had experienced discrimination at least once since they arrived (compared with 56% of the Angolans, 36% of the Cape Verdeans and 46% of nationals from other Portuguese-speaking African Countries). However, Mendes and Candeias (2013) did not access how frequently those experiences occurred. In addition, while Santos and colleagues (2005) asked participants about experiences with several behaviors that may be considered discrimination, Mendes and Candeias (2013) asked participants a single question about experiences with "discrimination," even though what constitutes discrimination may vary considerably among respondents. Finally, it should be highlighted that a considerable amount of the Brazilian population in Portugal can be categorized as Black, that is, as having an African origin (European Union Agency for Fundamental Rights, 2009).

percentages are based on a rather small number of events). It should be noted that the prevalence of these comments is likely higher, because the cited studies did not include expressions of bias that occurred in absence of members of the target group. In a more recent study, a mixed sample of White and non-White students reported witnessing an average of almost nine comments a week targeting other ethnic groups (Dickter & Newton, 2013).

Verbal expressions of prejudice are typically made by members of high-status groups about a member of low-status or traditionally disadvantaged group (Dickter, 2012; Dickter & Newton, 2013; Ayres, Friedman, & Leaper, 2009; Swim et al, 2003)[5]. Some groups tend to witness more prejudiced comments made by familiar people than prejudiced comments made by strangers. For instance, when women were asked to recall a time when they heard a sexist remark, about 20% of the described comments were made by strangers, while about 70% of the comments were made by a familiar person. Of this, about 60% were made by a person of the same status, such as a friend, a peer, or a co-worker (Ayres, et al, 2009). Similarly, studies using recall and diary methodology showed that most racially biased comments witnessed by college students (who were not the targets of that racial prejudice) were made by friends or acquaintances (Dickter & Newton, 2013; it should be noted, however, that Black people report more frequently being the target of prejudiced behaviors from strangers than from friends or acquaintances; Swim et al., 2003). These studies emphasize the importance of studying people's responses to biased comments made not only by strangers (as has been done more frequently in the past; e.g., Shelton & Stewart, 2004; Swim & Hyers, 1999; Woodzicka & LaFrance, 2001) but also by familiar people, as those episodes may be more frequent or salient.

Verbal expressions of prejudice perpetuate negative feelings and beliefs against the targeted groups (Blanchard, Crandall, Brigham, & Vaughn, 1994; Blanchard, Lilly, & Vaughn, 1991). Importantly, being the target of discrimination has, serious negative consequences for members of the targeted groups, decreasing their physical and mental health and wellbeing (for recent meta-analysis, see Paradies et al., 2015; Pascoe & Richman, 2009; Schmitt, Branscombe, Postmes & Garcia, 2014; for a study specifically

[5] Potentially biased action performed by a member of a dominant group may also be more likely to be detected and reported as prejudiced because members of dominant groups constitute the prototype of perpetrators of prejudice (Inman & Baron, 1996).

investigating the relation between experiences of discrimination and the well-being of immigrant adolescents residing in Portugal see Neto, 2001; 2006) Given these negative consequences, it is fundamental to investigate possible factors that facilitate or inhibit expressions of prejudice targeting racialized groups (that is, groups whose boundaries are defined by phenotypical characteristics commonly associated with "races;" Cabecinhas, 2002). One of the strategies may be to directly confront people expressing prejudice.

Confrontation: Addressing expressions of prejudice

Since expressions of prejudice are still prevalent today, people may face the decision of how to respond to those expressions, both if they are its targets or witnesses. People must choose whether to show agreement, to ignore the expression of prejudice or to confront prejudice, that is to "verbally or nonverbally expressing one's dissatisfaction with prejudicial and discriminatory treatment to the person who is responsible for the remark or behavior" (Shelton, Richeson, Salvatore, & Hill, 2006, p. 67; see Kaiser & Miller, 2004, for a similar definition)[6]. This definition allows multiple signals of disagreement to be considered confrontation of prejudice, as long as they are expressed directly to the person being biased. Thus, it highlights that confrontations can have different degrees of assertiveness, ranging from non-verbal expressions, such as facial expressions denoting disgust, to explicit, verbal disagreement with the comment (see, for example, Hyers, 2007; Shelton & Stewart, 2004; Swim & Hyers, 1999; Woodzicka & LaFrance, 2001).

Confrontations of bias address a socially unfair treatment and can be motivated by individuals' desire to promote egalitarian values. People who are confronted about their bias tend to refrain from expressing prejudice in the future (Czopp, Monteith & Mark, 2006). In addition, after being confronted with an accusation of prejudice, people may engage in efforts to mend their own self-image as an unprejudiced person and to repair the relationship with the confronter (efforts that are effective in increasing mutual liking;

[6] It is also pertinent to distinguish between confrontation of prejudice and attributions to discrimination. Even though they may sometimes overlap, attributions of discrimination do not always constitute confrontations of prejudice. Attributions to discrimination requires perceiving discrimination as the cause of an outcome and can be done publicly (to the person responsible for that outcome or to other people) or privately (see Kaiser & Miller, 2001; 2003; Garcia, Reser, Amo, Redersdorff & Branscombe, 2005). Confrontations of prejudice require perceiving a behavior, regardless of its outcomes, as motivated by prejudice, and that dissatisfaction with that prejudiced behavior is expressed directly to the person responsible for the behavior (see Kaiser, Hagiwara, Malahy, & Wilkins, 2009; Shelton et al., 2006).

Mallett & Wagner, 2011), especially if they are internally or externally motivated to respond without prejudice (Plant & Devine, 1998, 2009). Although some evidence suggests that both confronters who belong to the group target by the prejudiced behavior (henceforth called "targets") and confronters who do not belong to that group (henceforth called "witnesses") are effective in reducing bias (Czopp et al., 2006), other studies support the idea that witnesses' confrontations may be more persuasive than target confrontations (Czopp & Monteith, 2003; Drury & Kaiser, 2014; Gulker, Mark & Monteith, 2013; Rasinski & Czopp, 2010). The higher effectiveness of witnesses' confrontation (compared to targets' confrontation) may be due to it being more surprising. People tend to process more a message that is inconsistent with the source's self or group interest, which may increase the persuasiveness of that message (Petty, Fleming, Priester, & Feinstein, 2001). Accordingly, in Gulker et al. (2013) study, participants who were confronted about their own prejudice against Black people perceived the confrontation to be more unexpected when the confronter was White (vs. Black) and perceived the White confronter less as a complainer (also consistent with findings showing that targets who act against their self-interest tend to be evaluated more positively; Eagly, Wood, & Chaiken, 1978). The extent the confrontation was evaluated as surprising and evaluations of the confronter predicted, in turn, the extent people were willing to acknowledge their own prejudice (Gulker, et al., 2013). Non-target confrontations are also more effective in decreasing the agreement of people who watch witnesses confront (henceforth called observers) with a biased behavior and in increasing perceptions of the confronted person as prejudiced (Rasinski & Czopp, 2010).

Despite the potential of confrontation to reduce expressions of prejudice (Czopp et al., 2006), only recently researchers have started to focus on the psychological mechanisms explaining targets and witnesses' responses to bias (for a recent special issue on the topic, see Becker, Zawadzki, & Shields, 2014). Studies specifically focusing on factors predicting witnesses' responses to expressions of prejudice are even sparser, despite the important practically and theoretically implications of the topic. Besides informing practices aiming to reduce prejudice, studying witnesses' confrontations of bias is theoretically interesting, because the behavior lays in the intersection between interpersonal, intragroup and intergroup processes. Indeed, witnesses' confrontations of prejudice are interpersonal interactions, often between high-status group members regarding low-status group members (Dickter, 2012; Dickter & Newton, 2013; Ayres, et

al., 2009; Swim et al, 2003). Thus, confronting prejudice not only requires managing interpersonal relations but also one's position as a member of a dominant group who challenges fellow ingroup members on behalf of outgroup members.

Investigating witnesses' confrontations of bias

It is possible to investigate witnesses' confrontations of prejudice from three main perspectives. The first perspective considers how individuals' evaluate witnesses' responses to prejudice. It is generally represented by the following question: do observers' who see a witness confront prejudice evaluate the behavior positively? The second perspective is focused on how society globally perceives responses to prejudice: is witnesses' confrontation of bias a prescriptively normative behavior, that is, a behavior that is valued, approved and perceived as desirable by society (Cialdini, Reno & Kallgren, 1990; Dubois, 2003)? Finally, the third perspective pertains how people react when they witness prejudice toward a low-status group – to what extend do they confront or, alternatively ignore those expressions? These perspectives should not be seen as opposed, but as complementary, because attitudes, social norms and behaviors likely have a reciprocal relation with one another (e.g., Ajzen, 1991; Bem, 1972; Cialdini, Kallgren & Reno, 1991; Fazio, 1990).

In the present thesis, we aimed to contribute to the current understanding of witnesses' responses to bias by integrating social factors (emerging from the interpersonal relationship between the person expressing prejudice and the witness of bias) and individual factors (related to individual differences in egalitarian values) in predicting not only individual attitudes regarding witnesses' confrontation of bias, but also the extent witnesses' are actually willing to confront prejudice. As Doise (1980; 1986) highlighted, any social complex behavior is likely the product of the influence of multiple factors, of the same and of different levels of analysis: individual factors, relationships at the interpersonal and intergroup levels, and societal norms, all potentially influence people's attitudes and social behaviors. Witnesses' confrontation of bias should be no exception. In order to fully understand why and when witnesses confront prejudice, and how confrontations of prejudice are evaluated, it is necessary to integrate factors at different, complementary levels of analysis.

Thus, in the current dissertation, we relied on social and individual factors previously identified by research on confrontation of bias or, more broadly, in intergroup

research, and integrated them to predict attitudes and behavior regarding witnesses' responses to expressions of prejudice. We also investigated the effect of social factors in predicting prescriptive norms regarding witnesses' confrontation of bias. However, because prescriptive norms reflect beliefs shared by a social group, rather than beliefs of a given individual, about a behavior, we did not expect differences in value-based motivations to be important predictors of prescriptive norms. We will next briefly summarize research regarding each of the variables included in our studies, starting with the social factors and later presenting the individual factors.

Social factors

One of the most powerful predictors of targets' and witnesses' confrontations of bias is the relationship between the potential confronter and the person expressing bias. People usually report being more willing to confront a close person than a stranger (Ayres, et al., 2009; CNN/ORC, 2015). Different reasons can account for this difference. On one hand, people tend to feel more comfortable discussing emotionally charged topic with close people than with strangers (Clark & Taraban, 1991) and reactions from a close person who is confronted may be more predictable and less threatening, compared with the reactions of a stranger, making confrontation less costly (as suggested by Ayres, et al., 2009). On the other hand, it is expectable that witnesses believe it is more prescriptively normative to confront a close person than a stranger. Often, social norms suggest that "people should mind their own business" (e.g., Schwartz, 1973), and witnesses may refrain to confront because they feel they are not directly implicated in the prejudiced comment.

However, behaviors that benefit an ingroup tend to be perceived as socially appropriate and justified (Pereira, Vala, & Leyens, 2009; Ramos, Pereira & Vala, 2016), especially when those actions benefits a group that is highly valued (Kaiser, et al, 2009). It is likely that people perceive they have more social legitimacy to act when they feel that the image of a highly-valued ingroup is threatened. Because expressions of prejudice tend to be perceived as an immoral behavior (Dickter, Kittel, & Gyurovski, 2012; Effron, Miller, & Monin, 2012; Rutland, Killen & Abrams, 2010), ingroup members who express prejudice should threaten the image of the ingroup (Brambilla, Sacchi, Pagliaro, & Ellemers, 2013). A friend who expresses bias should threaten the image of a highly-valued group (i.e., the group of friends; Lickel, et al, 2000) more than a stranger who

expresses bias; and the more the image of a highly-valued ingroup is threatened the more people may perceive it is prescriptively normative to confront bias, to the extent that confronting would restore the ingroup image.

The extent the person expressing prejudice has power (i.e., the ability of control resources or outcomes of another individual; Keltner, Gruenfeld, & Anderson, 2003) over the potential confronter seems to be an important predictor of both target and witnesses' responses to prejudice. Women are less likely to confront sexism if the person expressing bias is someone who has power over them (e.g., if that person is a professor or a boss), compared with when the person has an equal status (e.g., is a friend or a co-worker; Ayres et al., 2009), presumably because confronting a person who has power over oneself implies potential higher costs. Similarly, women interviewing for a job were less likely to confront a sexist interviewer when confronting implied higher costs (i.e., they were being interviewed for a highly desirable job) than when confronting implied lower costs (i.e., the interview was for an undesirable job; Shelton & Stewart, 2004). Indeed, the perceived costs (vs. benefits) of confronting prejudice influence both targets' and witnesses' decision to confront (Ashburn-Nardo, Blanchar, Petersson, Morris, & Goodwin, 2014; Good, Moss-Racusin & Sanchez, 2012; Kaiser & Miller, 2004). Like targets, witnesses of prejudice are less willing to confront a boss than a co-worker, both because they feel less responsible to confront and because they anticipate more negative consequences of confronting someone who has power over them (Ashburn-Nardo, et al., 2014).

Besides influencing potential confronters' responses to prejudice, the costs of confronting may also influence observers' evaluations of those responses. Past research has shown that prosocial actions are evaluated less positively when those actions imply higher personal risks (Holahan, 1977; Piliavin, Dovidio, Gaertner, & Clark, 1981). Because witnesses' confrontational behaviors can be conceptualized as prosocial behaviors (as they benefit a person other than the self; Penner, Dovidio, Piliavin, & Schroeder, 2005), the costs of confronting may reduce the perceived appropriateness of a confrontational behavior and increase the acceptability of a non-confrontational behavior. However, we were not aware of any research investigating whether this is the case.

Decisions to confront or not confront may also be influenced by the extent people perceive the person expressing prejudice as an ingroup member or an outgroup member. People generally tend to perceive ingroup members as more moral, believing them to be

more trustworthy and honest (Brewer, 1999; Hewstone, Rubin, & Willis, 2002; Mullen, Brown, & Smith, 1992). Moreover, negative behaviors tend to be attributed less to dispositions of a person when that person is an ingroup member, compared with when that person is an outgroup member (Hewstone, 1990; Pettigrew, 1979; Weber, 1994; see also research on the linguist intergroup bias, Maass, Salvi, Arcuri, & Semin, 1989). These findings suggest that the group membership of the person expressing prejudice (i.e., whether that person is an ingroup or an outgroup member) may alter the way he or she is perceived by witnesses: they may attribute characteristics associated with racism (see Sommers & Norton, 2006) more strongly to an outgroup member who expresses bias than an ingroup member who performs the same behavior.

Because confrontations of prejudice are predicted by the extent a comment is perceived to reveal internal biases of the actor who makes the comment (Lee, Soto, Swim, & Bernstein, 2012), the attribution of characteristics associated with racism to a person expressing bias may, in turn, predict witnesses' willingness to confront. While the group membership of the person expressing prejudice may have been a factor influencing witnesses' reactions to expressions of prejudice in previous studies (e.g., in studies using a diary methodology where participants report hearing prejudiced comments from people from different groups; Dickter & Newton, 2013), so far no research has considered whether a shared group membership with the person expressing bias reduces the extent witnesses attribute prejudice-related characteristics to him or her and, consequently, the extent they are willing to confront.

Individual factors

At least in societies with anti-prejudice norms, confrontations of bias by witnesses tend to be perceived in a positive light (Dickter, et al, 2012), because they promote egalitarian values and address a behavior that is considered socially unjust. However, people differ in the extent they adhere to and have internalized egalitarian values (Plant & Devine, 1998, 2009; Schwartz, 1992; Schwartz et al., 2012). These individual differences likely predict observers' evaluations of responses to prejudice and witnesses' actual responses to prejudice.

Values can be defined as beliefs that guide not only a person's own actions but also the evaluation of other people's behaviors across situations (Schwartz, 1992). According to the theory of basic human values (Schwartz, 1992; Schwartz, et al., 2012), it is possible

to identify a set of values that is shared across cultures; however, people vary on how much they prioritize each value in that set. The theory of basic human values has been tested by cross-cultural research in more than 80 countries from different continents, including Portugal and the US, with results asserting its comprehensiveness and applicability (Schwartz, 2012; Schwartz, et al., 2001).

One of the proposed values of the theory of basic human values that is of special interest in the present thesis is the value of universalism (Schwartz, 1992; Schwartz et al., 2012). Universalism is defined as a motivation to understand, appreciate, tolerate, and protect all people and nature. Universalism is closely and positively related to benevolence (together, they constitute the supra-ordinate value of self-transcendence). However, while benevolence is defined as a motivation to protect and enhance the welfare of close-people, universalism is defined as a concern about all people, regardless of their relationship with the self (Schwartz, 1992; 2012). Accordingly, although both universalism and benevolence predict prosocial behaviors (Caprara, Alessandri & Eisenberg, 2012; Caprara & Steca, 2007), universalism is, conceptually, a better predictor of behaviors that benefit groups of people or people more distant from the self (Schwartz, 2010). Furthermore, universalism is usually more strongly (negatively) related to measures of prejudice and social dominance than benevolence (Cohrs, Moschner, Maes & Kielmann, 2005; Duriez & Van Hiel, 2002; Feather & McKee, 2008), and is an important predictor of attitudes toward immigrants and racialized groups (Peral & Ramos, 2014; Ramos, 2011; Ramos, Pereira & Vala, 2016).

Universalism can be further specified into three lower-order values motivations: universalism-nature, a motivation to preserve the natural environment; universalism-tolerance, a motivation to accept and understand people who are different from oneself; and universalism-concern, that represents a commitment to fight for equality and social justice (Schwartz et al., 2012). Universalism-concern is the value that is more closely related to attitudes toward other groups that are in a disadvantaged position in society. For instance, universalism-concern predicts attitudes toward homosexuals better than universalism-nature, and attitudes toward economic inequality better than universalism-tolerance (Schwartz et al., 2012). Universalism-concern is also more closely related to behaviors supporting a group fighting for equality than universalism-nature or universalism-tolerance (Schwartz & Butenko, 2014).

The degree to which observers endorse universalism-concern should determine their evaluations of witnesses' responses to bias, because values influence not only individuals' behaviors, but also their attitudes toward other people's behaviors (Homer & Kahle, 1988; Schwartz, 1992). Individuals who prioritize universalism-concern relatively to other values may evaluate witnesses' confrontations of prejudice more positively, as confrontations of prejudice directly affirm egalitarian values. However, no research had investigated this possibility.

Because values express individuals' goals and motivational concerns (Schwartz, 1992), it is possible to assess the extent people have internalized egalitarian values through their internal motivation to act without prejudice towards other people (Plant & Devine, 1998, 2009). Plant and Devine (1998) distinguish between two sources of motivation to respond without prejudice: an internal motivation to respond without prejudice, reflecting the extent non-prejudiced beliefs are important personal standards of conduct; and an external motivation to respond without prejudice, which reflects individuals' motivation to control their own prejudiced responses due to normative concerns and fear of social sanctions (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Plant & Devine, 1998, 2009).

Individuals who are internally motivated to respond without prejudice tend to act unbiasedly, expressing low levels of prejudice when answering to explicit measures of prejudice privately and publicly. A person who is externally, but not internally motivated to respond without prejudice, in contrast, tend to only act unbiased in the presence of others who may criticize their behavior (Plant & Devine, 1998; Plant, Devine & Brazy, 2003). While external motivation to respond without prejudice does not influence responses in explicit measures of prejudice for people who are internally motivated to respond without prejudice (Plant et al., 2003), the same is not true for responses that are harder to control, such as responses in implicit measures of prejudice. In such measures, individuals high in internal and low in external motivation to respond without prejudice tend to show lower levels of prejudice than those who are high both in internal and in external motivation and those who are not internally motivated to respond without prejudice (Devine et al., 2002).

Importantly, internal motivation to respond without prejudice has also been shown to predict people's distress after witnessing another person expressing bias. Participants higher in internal motivation to respond without prejudice revealed more physiological

reactions linked with negative affect and reported more negative emotions when watching an ingroup member expressing biased attitudes, compared with participants lower in that motivation. External motivation to respond without prejudice, on the contrary, was unrelated to both measures of negative affect (Schmader, Croft, Scarnier, Lickel, & Mendes, 2012). These results are not surprising, as external motivation to respond without prejudice does not reflect the internalization of egalitarian values, which constitute standards against which the behavior of others can be evaluated. Instead, it captures people's willingness to regulate their own bias according to an external norm or standard (Plant & Devine, 1998, 2009; Schmader, et al., 2012).

Based on Schmader et al. (2012) results, we hypothesized that people who are more internally motivated to respond without prejudice would be more willing to confront a person expressing prejudice because the stronger negative feelings they experience after witnessing another person expressing bias would transcend group membership. Because they have internalized strong egalitarian standards, people higher in internal motivation to respond without prejudice are more motivated to control their own prejudice toward other groups (Monteith, 1993; Plant & Devine, 1998); similarly, they may also be more motivated to reduce other people's prejudiced behavior. Past research on witnesses' confrontation of bias has focused on differences in attitudes toward different groups who are the target of prejudice, either showing that participants who have more favorable attitudes toward those groups confront prejudice more or showing a null relationship between attitudes and confrontation of prejudice (Dickter, 2012; Dickter & Newton, 2013). However, because internal motivation to respond without prejudice was shown to be both a predictor of explicit and implicit attitudes toward other groups (Devine et al., 2002) and, especially important, a predictor of reactions to other people's prejudice (Schmader et al., 2012), it may be a particularly relevant individual factor influencing witnesses' direct responses to a person expressing prejudice.

In summary, many factors (social, individual, and also societal) should influence people's behaviors and attitudes in any given situation (Doise, 1980, 1986). While previous research has examined several factors influencing both observers' attitudes toward confrontations of prejudice and targets' and witnesses' actual responses to prejudice, to our knowledge no studies have investigated the role of witnesses' and observers' endorsement and internalization of egalitarian values on those processes.

Furthermore, no studies have integrated social factors emerging from the relationship between the person expressing bias and the potential confronter with such value-based individual differences. Therefore, one of the contributions of the present thesis is the integration of factors from different levels of analysis to contribute to the understanding of witnesses' responses to expressions of prejudice and of observers' evaluations of those responses.

Overview of the current thesis

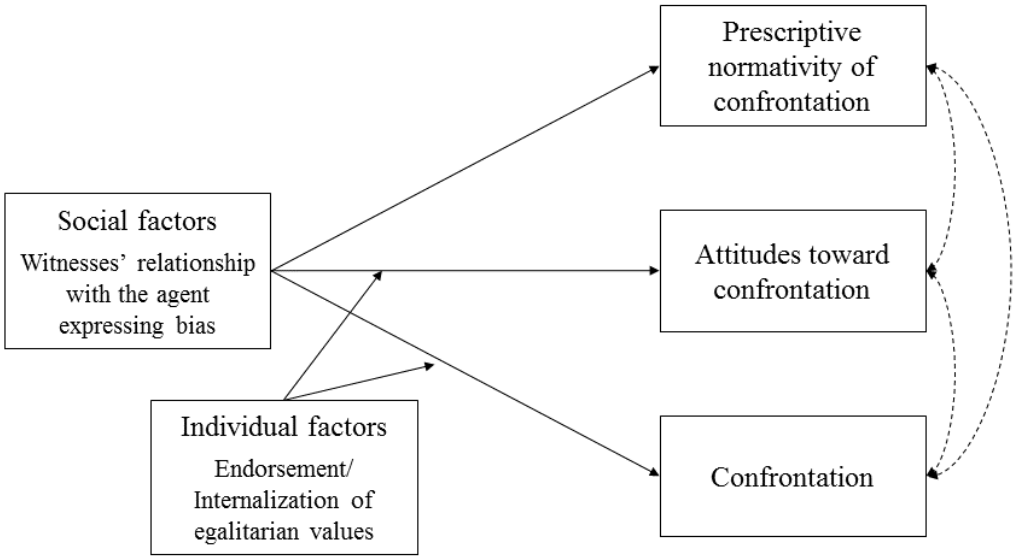
The current thesis investigated social and individual factors influencing witnesses' confrontation of bias[7] from three distinct perspectives: a) observers' attitudes toward witnesses' confrontation (and non-confrontation) of bias; b) witnesses' actual confrontation of bias; and c) the extend confrontation is perceived to be prescriptively normative. More specifically, it was hypothesized that characteristics of the relationship between the potential confronter and the person expressing bias would predict each of the three outcomes (norms and attitudes regarding confrontation of bias, and behavioral responses to bias). We hypothesized that the influence of variables related with the relationship between the potential confronter and the person expressing bias on attitudes toward responses to bias and actual responses to bias would depend on witnesses' and observers' endorsement and internalization of egalitarian values, but we expected prescriptive norms to be relatively unaffected by individual differences in value-based motivations, because prescriptive norms reflect rules and standards shared by members of social groups (Cialdini et al., 1991; Cialdini & Trost, 1998).

The proposed relations between the social and individual variables investigated in this dissertation are represented in Figure 1. It should be noted that although it is likely that norms, attitudes and behavior influence one another, we did not test the actual relationship among these variables. For that reason, we represented the relation among

[7] Even though in the current introduction we use both the terms "bias" and "prejudice", in the empirical chapters of this thesis we opted to use the term "expression of bias" instead of "expression of prejudice" or "expression of stereotypes." While prejudice can be defined as negative affect toward people due to their membership in that group (Brown & Lepore, 1996) and stereotypes can be defined as shared beliefs about the characteristics that are perceived to be true of members of particular social groups (Stangor, 1996), bias is a more encompassing term that may include both expressions of prejudice and of stereotypes (Dovidio & Gaertner, 2010). Indeed, expressions of bias against other groups often include a mix of expressions of negative affect and negative beliefs about the characteristics of those groups.

prescriptive normativity of confrontation, attitudes toward confrontation, and actual confrontation of prejudiced with two-way arrows and dashed lines.

Figure 1. Proposed relations between social and individual factors and the normativity of witnesses' confrontation of prejudice, attitudes toward witnesses' confrontation of prejudice and actual confrontations of prejudice by witnesses.



To test the hypothesized relations among these variables, seven studies were conducted. These studies are presented in three empirical chapters, written as individual papers.

In Chapter II, we showed that witnesses' confrontations of bias are generally more positively evaluated than non-confrontations of bias, in line in previous research (Dickter, et al., 2012). This was specially the case when the relation between the potential confronter and the person expressing bias implied that confronting would have low costs (i.e., the person expressing bias was an interviewer for a job the confronter does not need). However, we further contributed to the literature by showing that when confronting would entail high costs (i.e., the actor expressing bias was an interviewer for a job the confronter really needs), confronting was not evaluated as a more appropriate behavior than non-confronting (Study 1, Chapter II). We argue that when the actor expressing bias was an interviewer for a needed job, the personal costs of confronting (i.e., losing a much

needed job opportunity) and the social costs of not confronting (i.e., not addressing an unfair behavior) were in conflict, increasing the ambiguity of the appropriateness of different responses to bias. In more ambiguous situations, individual factors become important determinants of evaluations of others' behaviors (Mischel, 1973; Snyder & Ickes, 1985). Therefore, in Study 2 of Chapter II, we focused on the high cost situation. Results showed that attitudes toward responses to bias depend on individuals' endorsement of universalism-concern, that is, on the extent participants are committed to fight for equality and social justice (Schwartz et al., 2012). In the high cost situation, only participants who endorse universalism-concern more strongly evaluated confrontation of bias as more appropriate than non-confrontation.

In Chapter III, we investigated the effect of a shared group membership with the person expressing bias in witnesses' willingness to confront bias, a factor previously overlooked in the confrontation literature. Because people tend to make more dispositional attribution for negative behaviors when the actor is a member of an outgroup than when the actor is a member of an ingroup (Hewstone, 1990; Pettigrew, 1979; Weber, 1994), we hypothesized (and found) that witnesses would be less willing to confront an ingroup than an outgroup member, and that these effect would be explained by witnesses attributing bias-related characteristics to the person expressing bias less strongly when that person was an ingroup (vs. an outgroup) member. However, this effect was moderated by the extent witnesses had internalized egalitarian values. Only participants lower in internal motivation to respond without prejudice (Plant & Devine, 1998) attributed bias-related characteristics less strongly to an ingroup member who expressed bias than to an outgroup member who expressed bias, and consequently, confronted the outgroup member more; we found no such difference for participants higher in internal motivation to respond without prejudice.

In Chapter IV, we hypothesized that the prescriptive normativity of witnesses' confrontations of bias would depend on who the person expressing bias is. More specifically, we predicted that it would be more normative to confront a close person (i.e., a friend or a relative) than a stranger. Confrontations of bias are sometimes perceived as impolite or out-of-place behaviors; however, one of the factors that may increase the normativity of confrontational behaviors is threat to the image of a highly-relevant ingroup, such as a group of friends. Confronting bias may restore the image of an ingroup, by reinforcing egalitarian norms, and behaviors that aim to protect an ingroup are

generally socially approved (Pereira et al, 2009; Ramos et al, 2016), especially when that group is highly-valued (Kaiser et al, 2009). Results supported the hypothesis. Confronting bias is generally prescriptively normative, but confronting a close person is more normative than confronting a stranger. Threat to the image of a highly-valued group mediated this effect. A close person who expresses bias threatened the image of a highly-value group more; and the more the image of a highly-valued group was threatened, the more was socially valued and approved to confront. These results suggest that threat to the image of a highly-relevant group grants witnesses social legitimacy to challenge expressions of prejudice, making confrontations more normative.

Finally, in Chapter V, the results of the empirical chapters are integrated, and the theoretically and practical contributions of the thesis are discussed. We also discuss the limitations of the present work and propose future direction for this line of research.

In summary, we investigated the joint influence of social and individual factors underlying three different facets of confrontation of bias: witnesses' actual responses to biased comments, observers' attitudes toward confrontation and the prescriptive normativity of witnesses' confrontations of bias. Our main contributions to the understanding of confrontations of bias resides in three features of the current thesis. First, we investigated witnesses' confrontations of bias, a topic understudied in the literature. Second, we investigated the influence of key variables that had been overlooked in previous research in witnesses' confrontation of bias, namely the group membership shared by the witness and the person expressing bias, and the threat to the image of a highly-valued ingroup. Third, we showed that the influence of social factors related to the relationship between the person expressing bias and the potential confronter on attitudes and behavioral responses to bias depends on observers' and witnesses' individual differences in egalitarian values. People who are strongly committed to equality and social justice confront and evaluate positively another person even in less favorable social conditions. Therefore, our findings directly contribute to the current understanding of witnesses' responses to bias, interpersonal behaviors that may fundamentally allow or prevent public expressions of prejudice.

Chapter II

Evaluations of Witnesses' Responses to Bias: Universalism-Concern and the Costs of Confrontation

This chapter is based on:

Lavado, S., Pereira, C. R., Dovidio, J. F., & Vala, J. (2016). Evaluations of witnesses' responses to bias : Universalism – Concern and the costs of confrontation. *Personality and Individual Differences, 96,* 172–180. doi:10.1016/j.paid.2016.03.003[8]

[8] As the first author, I was the primary responsible for the design, implementation, data analysis and writing of the studies reported in this paper. My co-authors (my doctoral advisors) provided fundamental guidance and supervision throughout the entire process, ultimately leading to the publication of this manuscript.

Abstract

The present research examined how situational and individual difference factors influence majority-group observers' evaluations of witnesses' responses to an incident of bias. In Study 1, participants learned of a situation in which a White person applying for a job he did or did not need (high vs. low cost of confrontation) heard his interviewer make a racist comment, which the White person did or did not confront. Non-confrontation was evaluated as less appropriate than confrontation when the costs of confronting were low, but not when costs were high, revealing that in a high cost situation the appropriate response to bias is more ambiguous. Study 2 focused on this high cost situation to show that evaluations of another person's responses to bias depend on individual differences in the observer's values. Observers who scored low on Universalism-Concern evaluated another person's non-confrontation as appropriate as confrontation, but participants who scored high on Universalism-Concern perceived non-confrontation as less appropriate. Considering how responses to bias are assessed helps illuminate normative processes that affect confrontations of bias against outgroups, contributing to the knowledge of the processes that may allow biases to persist.

Key-words: bias, confrontation, costs, universalism-concern, witnesses

Introduction

Recent research has aimed to understand when targets of bias confront unfair negative comments and actions directed toward them or their group (see, for example, Becker, Zawadzki, & Shields, 2014). That research also considers how individual differences, such as in beliefs about the malleability of prejudice (Rattan & Dweck, 2010) or optimism (Kaiser & Miller, 2004), among targets can increase or decrease their willingness to confront bias. However, confronting bias is not solely the responsibility of members of targeted, disadvantaged groups; how members of majority groups not only perceive injustice (Drury & Kaiser, 2014; Inman & Baron, 1996) but also evaluate the responses of others to injustice can affect the persistence and impact of social bias in society. In the present research, consisting of two studies, we investigated how majority-group members evaluate other ingroup members who do or do not confront racial bias against another group. Specifically, we tested the potential moderating roles of (a) the social conditions under which the person decided whether to confront the bias (Study 1), and (b) individual differences in the values held by observers of the other person's response to bias (Study 2).

Background

Confrontations of bias address socially unfair treatment and preserve egalitarian norms and are therefore generally seen by observers as positive social behaviors (Dickter, Kittel, & Gyurovski, 2012). Non-confrontations of bias are typically perceived as less appropriate, because they allow a biased remark to remain unchallenged, and may even convey agreement with bias.

Costs of Confrontation. Although confronting bias may generally be perceived as a socially responsible act, how people evaluate the appropriateness of confrontation and non-confrontation may be shaped by contextual influences. Previous research on prosocial behavior has highlighted that perceiving personal costs associated with performing the act affects the way people evaluate the appropriateness of both engaging in and refraining from enacting the prosocial behavior. For instance, when helping involves greater personal cost to the person who intervenes (e.g., greater personal risk), not intervening is perceived to be a more socially acceptable response (Holahan, 1977; Piliavin, Dovidio, Gaertner, & Clark, 1981). Indeed, people often justify not intervening

to help another person on the basis of the potential costs incurred for helping (Penner, Dovidio, Piliavin, & Schroeder, 2005).

Confrontation of bias by a witness may also be considered a form of prosocial behavior involving assessments of costs and benefits of various courses of action (Ashburn-Nardo, Blanchar, Petersson, Morris, & Goodwin, 2014; Penner et al., 2005). As such, when people recognize that the personal costs are high to a witness for confronting bias, they may perceive non-confrontation as excusable, and thus as relatively socially appropriate. We tested this general hypothesis in Study 1.

Values and Appropriateness of Confrontation and Non-Confrontation. The degree to which people evaluate the appropriateness of confrontation or non-confrontation by a witness may also be affected by relevant individual differences, particularly in situations in which the personal costs of confronting are at odds with the social costs of *not* confronting bias. In situations in which behavioral appropriateness is ambiguous or there are conflicting influences, individual differences among observers are particularly important guides of behavior (e.g., Eccleston & Major, 2006). As Mischel (1973) explained, "Individual differences can determine behavior most strongly when the situation is ambiguously structured … so that subjects are uncertain about how to categorize it" (p. 276; see also Snyder & Ickes, 1985). Previous research has revealed that confrontations by individual targets of bias are shaped by personal factors, such as commitment to fight bias or optimism (Kaiser & Miller, 2004; Shelton, Richeson, Salvatore, & Hill, 2006). However, we are not aware of research on individual differences in how people assess the appropriateness of others' confrontation or non-confrontation of bias.

Observers' values likely influence their judgments of the appropriateness of another person's action or inaction in the face of bias. Values are general beliefs that guide not only people's selection of actions but also evaluations of their own and other people's behaviors (Feather, 1995; Rokeach, 1973; Schwartz, 1992; 1994; Schwartz & Bilsky, 1987), particularly members of their own group's (Marques & Paez, 1994; Tyler & Blader, 2003). Values would likely be important predictors of the weight given to different costs and rewards in responses to bias, because values directly define the standards by which action (or inaction) is judged.

Schwartz's (1992) original theory of basic human values attempted to capture a comprehensive and cross-culturally valid set of values and to describe the relations among

those values. Schwartz identified 10 basic human values (Self-Direction, Stimulation, Hedonism, Achievement, Power, Security, Conformity, Tradition, Benevolence and Universalism) that can be organized into a circular continuum, according to compatibilities and conflicts among them. Cross-cultural research in more than 80 countries and with diverse samples supported the comprehensiveness of this set of values, their relationships, and their broad applicability (see Schwartz, 1992, 2012; Schwartz, Melech, Lehmann, Burgess, Harris, & Owens, 2001). The theory of basic human values has been used in research on diverse topics, such as political behavior (Schwartz, Caprara, & Vecchione, 2010), self-affirmation (Burson, Crocker & Mischkowski, 2012), and altruism (Lönnqvist, Leikas, Paunonen, Nissinen, & Verkasalo, 2006).

Universalism, the value of primary interest in Study 2, represents a motivation to understand, appreciate, tolerate, and protect all people and nature. Universalism is closely (and positively) related to Benevolence. However, Benevolence is defined as a motivation to care for the welfare of people with whom one is close and therefore has a relatively narrow focus of application. By contrast, Universalism is related to concerns about the welfare of others more generally. Both Universalism and Benevolence are in conflict with Power (a motivation to attain social status and prestige, and control or dominance over people and resources) and Achievement (a motivation to be personally successful according to social standards) (Schwartz, 1992).

While Universalism, Benevolence, Achievement, and Power are all related to traditional measures of social bias, Universalism is the value most strongly related to measures of prejudice and social dominance (Cohrs, Moschner, Maes, & Kielmann, 2005; Duriez & Van Hiel, 2002; Feather & McKee, 2008) – individuals who more strongly endorse the value of Universalism score lower on these measures. In addition, although Benevolence and Universalism (but not other values in the model) are important predictors of prosocial behaviors (Caprara, Alessandri, & Eisenberg, 2012; Caprara & Steca, 2007), Universalism is more closely related conceptually to prosocial actions toward other people in general, not just toward others with whom one is close (Schwartz, 2010). Because values affect behavior mainly when they are activated by a specific situation (Verplanken & Holland, 2002) and the value of Universalism captures whether equality is held as a central standard of behavior, we hypothesized that Universalism would be the primary value in guiding evaluations of confrontations (and non-confrontations) of bias.

In addition, Schwartz and colleagues (2012) recently refined the theory of basic human values and identified three subtypes of Universalism: Universalism-Nature, a motivation to preserve the natural environment; Universalism-Tolerance, a motivation to accept and understand people who are different from oneself; and Universalism-Concern a "commitment to equality, justice and protection for all people" (Schwartz et al., 2012, p. 669).

To the extent that Universalism-Concern specifically reflects a motivation to strive for social justice and equality, even at personal expense, when appraising the appropriateness of different responses to bias, people relatively high on this value would likely give more weight to the social cost of not confronting, even when there are potentially mitigating personal costs associated with confronting. In Study 2, we apply the situation identified in Study 1, where the personal costs of confronting are at odds with the social costs of not confronting, to test the unique effects of Universalism-Concern over and above other basic values in Schwartz et al.'s (2012) refined theory on observers' assessments of the appropriateness of non-confrontation versus confrontation of bias.

Purpose

The purpose of the present research was to examine how situational factors (Study 1) and individual differences in values (Study 2) influence majority-group observers' evaluations of witnesses' responses to an incident of bias. The aim of Study 1 was to understand how observers assess the appropriateness of not confronting (vs. confronting) as a function of situational factors affecting personal costs for intervention. The goal of Study 2 was to illuminate how individual differences in endorsement of a value related to the degree to which equality is held as a central standard of behavior (i.e., Universalism-Concern) influence observers' evaluations of different responses to bias in situations in which the cost to a witness for confronting bias is high.

Theoretically, expanding the study of confrontation to how others evaluate those who do or do not confront bias can broaden the perspective on the general social forces that can either ameliorate or maintain social bias. Practically, understanding the influences on non-targets who witness bias can have important social consequences, as non-targets who confront are taken more seriously and are seen as more persuasive than confronters who are the target of bias (Czopp & Monteith, 2003; Gulker, Mark, &

Monteith, 2013; Rasinski & Czopp, 2010). Investigating how observers evaluate witnesses' decision to confront bias or not and the conditions that may shape that assessment can provide insight into the process that socially inhibit unfair bias (Blanz, Mummendey, & Otten, 1997).

Study 1

Overview

In Study 1, participants (all from a majority group) learned of a situation in which a White applicant heard his interviewer make a biased comment about Black applicants and then did or did not confront the interviewer about that comment. We also varied the social circumstances of the applicant by indicating that he had a high versus low need for the position for which he was interviewing. The dependent measure was how appropriate participants perceived the behavior of the applicant.

Because confronting bias tends to be seen as a prosocial behavior that preserves egalitarian norms, we expected that a White witness confronting an expression of bias against a Black person would generally be seen as more appropriate social behavior than would non-confrontation. However, we further hypothesized that this effect would be diminished when the personal costs to the witness for confronting bias were relatively high (i.e., the applicant had a high vs. low need for the job). The costs of confronting would make non-confrontation more excusable and, thus, more appropriate.

Both studies in the present paper were conducted in the context of race relations in Portugal. Previous research has suggested that discrimination against Black immigrants in Portugal is generally condemned by social norms (Vala, Lopes, & Lima, 2008; Vala & Pereira, 2012). Concurrently, however, Black immigrants still report being the target of verbal harassment more often than other immigrant or ethnic groups (Santos, Oliveira, Kumar, Rosário, & Brigadeiro, 2009). Thus, we expected that participants would find the situation presented in these studies plausible and would consider the blatantly biased comment as unfair and inappropriate.

Method

Participants. With the permission of the university, during a week in mid-November, 2013, a female experimenter approached students in classroom locations across the campus and asked them to take part voluntarily in a study about people's

opinions about others' behaviors. Approximately 91% of the students approached consented to participate. The surveys, which reflected the different experimental conditions, were arranged randomly, and the experimenter was unaware of the condition a participant received. Participants received no direct compensation for their participation in the study.

Based on a priori power estimates for detecting an effect of moderate strength (G*Power; Faul, Erdfelder, Lang, & Buchner, 2007), to achieve power of .80 the targeted sample was 90 participants. A total of 106 students agreed participate. Because the theoretical focus in the research was on responses of members of the majority racial group as a function of whether another member of their group confronted bias against a racial minority group, Portuguese undergraduate students (n = 87; 55 men, 31 women, 1 did not specify gender; mean age = 20.69 years, SD = 4.05; participants' ages ranged from 18 to 41 years) were included in the analyses. An additional 19 participants who completed the survey but indicated a different nationality were not included in the final analyses.

The protocol for the research (for both Study 1 and Study 2) was approved by the Institutional Review Board (IRB) of the university where one of the authors is affiliated, and was consistent with the ethical standards and procedures of the university in Portugal. No information personally identifying participants was collected. Study materials and data (for both Study 1 and Study 2) are available upon request from the first author.

Design and procedure. Study 1 employed a 2 (Need for the Job: High vs. Low) x 2 (Behavior: Confrontation vs. Non-Confrontation) between-groups factorial design. Similar to the procedure of Shelton and Stewart (2004, Study 1), participants were presented with a scenario in which a candidate was being interviewed for a job. In this scenario, the interviewer indicated that he was favorably impressed by interviewee Paulo (a name selected because it is one of the most common names for White Portuguese men), but followed that with a racist comment about Black applicants for the same position. The potential cost to the interviewee for confronting biased remark was varied by information in the scenario suggesting that the interviewee had a high need for the job (high cost for confrontation) or a low need for the position (low cost of confrontation). Specifically, the participants read the following text (low need for the job condition in brackets):

> Paulo is in a room waiting to be called for a job interview. This interview is [not] very important to Paulo because this is his third job interview in months

> and no one has offered him a job yet [he already received some interesting job offers]. Paulo [doesn't want] wants to give his best shot at this interview because he [doesn't] really needs the job. Plus, this position [does not seem] seems really interesting and he [doesn't want the opportunity to work in the company that much] would love the opportunity to work in the company. Paulo is called in to the interview. He is greeted by the interviewer, a tall White man wearing a suit. They both sit down and he starts asking Paulo questions. Paulo has the impression that the interview is going well. In the end, the interviewer shakes his hand and says: I really liked you and I think you and the company would be a good fit. We had a lot of Black applicants, so it's good to have someone White for a change. I'll contact you when we have a decision.

We then varied the description of the behavior after the interviewer's racist comment. In the Confrontation condition, participants were told that the interviewee answered, "I don't think skin color should have anything to do with this." In the Non-confrontation condition, participants were told that the interviewee stated simply, "I'll be waiting for your call."

After participants read the scenario, they completed items measuring their perceptions of the appropriateness of the interviewee's behavior during the interview. In particular, participants were asked to evaluate, from 1 = *not at all* to 7 = *very much*, the degree to which the interviewee's behavior reflected four qualities: appropriate, wise, and (reverse-coded) irrational and unreasonable. The responses to the four items were averaged to form a behavioral appropriateness scale, $\alpha = .79$.

To evaluate the effectiveness of the manipulations, we also asked participants to rate, on separate 1 = *not at all* to 7 = *very much* scales, how important it was for him to get the job and the likelihood that he would be offered the position. Finally, we asked participants for their age, gender and nationality.

Results

We began by checking the normality of the distribution of the residuals on our main variable (behavioral appropriateness), which was verified, Shapiro-Wilk(87) = .981, p = .223, skewness of -.66 (SE = 0.26) and kurtosis of -.04 (SE = 0.51). Because our predictors were both categorical variables, we analyzed our data using a 2 (Need for the Job: High

vs. Low) x 2 (Behavior: Confrontation vs. Non-Confrontation) analysis of variance (ANOVA).

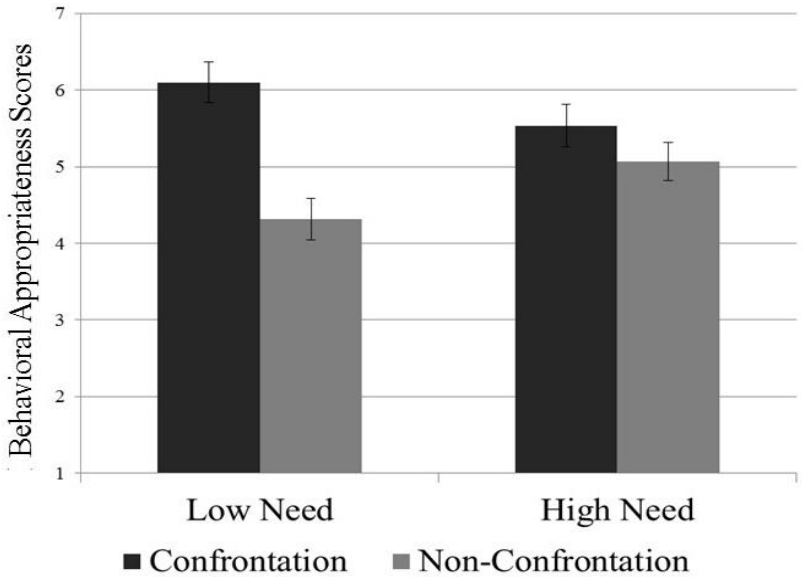
The manipulation produced the intended effects on the perceptions of participants. A 2 (Need for the Job: High vs. Low) x 2 (Behavior: Confrontation vs. Non-Confrontation) ANOVA yielded a main effect of need for the job on the importance to the interviewee of getting the job, $F(1,83) = 939.13$, $p < .001$, $\eta^2_p = .92$. Participants in the high need for the job condition perceived that it was more important for the interviewee to get the job ($M = 6.87$, $SD = .40$) than participants in the low need for the job condition ($M = 2.05$, $SD = .96$). There was no effect for Behavior ($F(1,83) = .02$, $p = .898$, $\eta^2_p < .01$) and no interaction effect ($F(1,83) = 1.17$, $p = .283$, $\eta^2_p = .01$). Also, as anticipated, participants perceived that the interviewee would incur personal costs for confronting the interviewer about the racist remark. A 2 x 2 ANOVA on the likelihood the interviewee would receive the job offer revealed a main effect for Behavior, $F(1,83) = 14.11$, $p < .001$, $\eta^2_p = .15$. Participants in the confrontation condition perceived that he would be less likely to receive the job offer ($M = 4.80$, $SD = 1.38$) than those in the non-confrontation condition ($M = 5.74$, $SD = 1.02$). There was also a marginally significant main effect of cost, $F(1,83) = 3.56$, $p = .063$, $\eta^2_p = .04$. Participants in the high need condition perceived it was less likely for the interviewee to get the job ($M = 5.09$, $SD = 1.40$) than participants in the high cost condition ($M = 5.52$, $SD = 1.13$). We found no interaction effect ($F(1,83) = .107$, $p = .745$, $\eta^2_p < .01$).

Addressing our primary research question, we conducted a 2 x 2 ANOVA on behavioral appropriateness. There was, as expected, a main effect of confrontation, $F(1,83) = 17.78$, $p < .001$, $\eta^2_p = .18$. Participants in the confrontation condition judged the behavior as more appropriate ($M = 5.83$, $SD = 1.14$) than participants in the non-confrontation condition ($M = 4.73$, $SD = 1.38$). Importantly, this effect was qualified by a significant interaction between Need and Behavior, $F(1,83) = 9.32$, $p = .016$, $\eta^2_p = .07$ (see Figure 1).

Planned comparisons revealed that when the costs were low, there was a significant difference between confronting and not confronting bias conditions, $F(1,83) = 21.71$, $p < .001$, $\eta^2_p = .21$. Confronting bias was evaluated as more appropriate ($M = 6.10$, $SD = .78$) than not confronting ($M = 4.32$, $SD = 1.50$). However, when the costs were high there was no difference between confronting bias ($M = 5.54$, $SD = 1.38$) and not confronting ($M = 5.07$, $SD = 1.20$), $F(1,83) = 1.58$, $p = .212$, $\eta^2_p = .02$. From an alternative perspective,

when the interviewee confronted the person interviewing him, his behavior was rated as equivalently socially appropriate whether his need for the job was high (and thus the costs for confrontation were high) or low (*M*s = 5.54 vs. 6.10), $F(1,83) = 2.13$, $p = .148$, $\eta^2_p = .03$. However, not confronting the comment was viewed as more socially appropriate when the interviewee's need for the job was high than when it was low (*M*s = 5.07 vs. 4.32), $F(1,83) = 4.16$, $p = .044$, $\eta^2_p = .05$.

Figure 1. Evaluations of behavioral appropriateness for each experimental condition of Study 1.



Discussion

The current study complements previous research, which shows that people are less likely to confront a biased remark when the costs of confronting are higher (Ashburn-Nardo et al., 2014; Shelton & Stewart, 2004), by investigating how *others* perceive the behavior of individuals in such situations. Consistent with previous research (Dickter et al., 2012), the results of Study 1 demonstrated that when an individual blatantly exhibits bias, participants perceived confrontation as a more socially appropriate response than non-confrontation. Future research might further investigate the processes contributing to this evaluation. For example, one reason why confrontation may be viewed so favorably is because such actions are positively distinctive: Majority-group members rarely respond to expressions of prejudice (Kawakami, Dunn, Karmali, & Dovidio, 2009). Another

reason why the type of confrontation represented in Study 1 may be viewed positively is because the bias was directed toward a group for which there are strong norms against biased treatment (Blacks), and the interviewee had no immediate self-interest in confronting the interviewer (Eagly, Wood, & Chaiken, 1978). It is possible that confronting a biased statement about another group would be perceived as less socially appropriate if the norms regarding prejudice against that group were less strong (e.g., overweight people) or possibly even support negative treatment (e.g., criminals) (Crandall, Eshleman, & O'Brien, 2002).

Study 1, however, further demonstrates that how people evaluate non-confrontation of bias depends upon their understanding of the circumstances of the other person who does not intervene. Importantly, Study 1 offers direct evidence of the moderating role of perceived mitigating circumstances on how inappropriate non-confrontation of even blatant bias is perceived. Specifically, although confrontation of bias was generally seen as more socially appropriate than non-confrontation of bias, majority-group observers tended to excuse non-confrontation. They rated non-confrontation as more socially appropriate (and as socially appropriate as confronting bias) when the costs to a witness of confronting bias are relatively high (i.e., not getting a job that was needed). Practically, the present results suggest that people who encounter an incident of bias may face different types of costs (or lack of rewards) for action or inaction, some emanating directly from the perpetrator (e.g., a boss) but others associated with the way observers evaluate their behavior.

Study 1 suggests majority-group observers are responsive to at least two different types of social forces in judging the appropriateness of confronting or not confronting expressions of racial bias. On the one hand, observers seem sensitive to prevailing egalitarian norms against racial bias and, as a consequence, evaluate a White person's confrontation of racial bias by another White person as more socially appropriate than a non-confrontation of such behavior. On the other hand, observers are responsive to the circumstances a witness of bias encounters in making a decision about whether to confront. Specifically, observers appear more willing to excuse a witness for not confronting bias when such action involves greater personal cost.

When these two factors – the cost to society for not intervening and the cost to an individual for confronting bias – are in conflict, the appropriate response to bias becomes more ambiguous. In this situation – our situation of interest in Study 2 – we hypothesized

the cost that would be more influential would relate systematically to individual differences between observers. We predicted that people who value social justice more would be more sensitive to the social costs of not confronting a behavior that violates egalitarian norms. We tested this hypothesis in Study 2.

Study 2

Overview

Study 2 examined whether the endorsement of Universalism-Concern predicts how majority-group observers evaluate a confrontation of bias versus a non-confrontation in a situation where the appropriate response to bias is ambiguous (i.e., the high need/high cost situation, identified in Study 1). We tested the unique effects of Universalism-Concern over and above the effects of other values, which complement or are in conflict with Universalism-Concern (Universalism-Tolerance, Universalism-Nature, Benevolence-Dependability, Benevolence-Caring, Achievement, Power-Dominance, and Power-Resources) in Schwartz et al.'s (2012) refined values theory. As noted earlier, the value of Universalism-Concern distinctively emphasizes the importance of equality in the treatment of all people, and this value is likely to be activated when people are exposed to an incident of injustice, such as bias (Verplanken & Holland, 2002).

We hypothesized that when the personal interests of a person who witnesses bias are in conflict with the general social interest in limiting expressions of bias, individual differences in observers' commitment to strive for equality would be important predictors of their evaluations of responses to bias. Specifically, we predicted that whereas participants lower in Universalism-Concern would perceive non-confrontation as appropriate as a confrontation, those higher in Universalism-Concern would perceive non-confrontation as less appropriate than confrontation, giving more weight to the social cost of not addressing an unfair treatment in their judgments.

Method

Participants. Based on the effect size and parameter estimates obtained in Study 1, a priori power analysis identified a target sample size of 110 participants to achieve the desired power of .80 in Study 2 (Faul et al., 2007). One hundred and twenty undergraduate students completed the study (60 men, 60 women; mean age = 21.84 years, SD = 3.61; participants' ages ranged from 18 to 33 years). A male and a female experimenter

recruited participants in a university library during two weeks in May/June 2014. All students present at the library at the recruitment time were individually approached and asked to complete the survey. Approximately 72% of the students who were asked agreed to participate. The study was introduced as an evaluation task, in which participants would be asked to give their opinion about another person's behavior in a social situation. The experimenters were unaware of the condition represented in the survey they asked participants to complete.

Design and Procedure. Participants were first asked to respond on a 6-point scale *(1 = not like me at all, 2 = not like me, 3 = a little like me, 4 = somewhat like me, 5 = like me, 6 = very much like me*) to the three Universalism-Concern items of the Schwartz et al. (2012) revised Portrait Values Questionnaire (PVQ-5X): "Here we briefly describe some people. Please read each description and think about how much each person is or is not like you:" "Protecting society's weak and vulnerable members is important to him[her]"; "He [she] thinks it is important that every person in the world have equal opportunities in life"; "He [she] wants everyone to be treated justly, even people he/she doesn't know." The scale showed good reliability in the current sample ($\alpha = .72$).

The items were embedded among items from the PVQ-5X measuring other related values: Universalism-Tolerance (motivation for accepting and understanding people who are different from oneself), Universalism-Nature (motivation for preserving nature), Benevolence-Dependability (motivation for being a dependable ingroup member), Benevolence-Caring (motivation for caring for other ingroup members), Achievement (motivation to succeed according to social standards), Power-Dominance (motivation for having power through control of other people), and Power-Resources (motivation for having power through control of material and social resources) (Schwartz et al., 2012). Because we believed the scenario would be mainly related to concerns about social justice, we did not have specific predictions for individual differences in these additional values. However, we included this items in order to investigate the unique predictive ability of Universalism-Concern, not only relative to other Universalism values but also, more broadly, to other self-transcendence and self-enhancement values.

Then, participants read in the survey the scenario representing the high cost context of Study 1. Participants were randomly assigned to one of two confrontation conditions varied in Study 1. In one condition the interviewee confronted the interviewer who made a racist comment; in the other condition the interviewee did not confront the interviewer.

After reading the scenario, participants answered the same Behavioral Appropriateness items (appropriate, wise, irrational and unreasonable) of Study 1 (α = .77). Then, participants were asked to indicate (from 1 = *not at all* to 7 = *very much*) their perceptions of the interviewer's behavior as (a) prejudiced, (b) appropriate, and (c) fair. Responses (reverse-coded for fair and appropriate) were averaged to produce a measure of perceived bias (α = .82). At the end of the survey we asked participants to recall whether the interviewee's had confronted or not confronted the prejudiced comment (as an attention check) and to provide information about their age, gender, and nationality[9]. All participants indicated they were Portuguese. We excluded from the final sample 11 participants who failed to correctly answer the attention check[10]. Thus, the final sample was composed by 109 participants.

Results

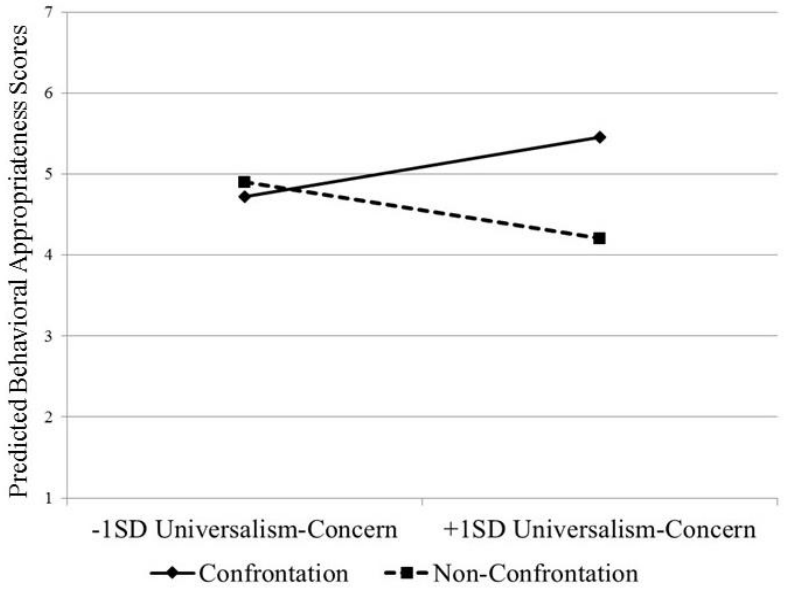
Because we had a categorical (Behavior: confront or not confront) and a continuous (Universalism-Concern) predictor, and a continuous dependent variable, we used linear regression analysis to test our hypothesis, which allowed us to test for the main effects and interaction of our variables of interest (Aiken & West, 1991), as well as to control for the effects of other variables. In order to do so, we began by checking the normality of the distribution of the residuals on our main variable dependent (behavioral appropriateness), which was verified, Shapiro-Wilk(109) = .986, p = .314, skewness of -.10 (SE = 0.23) and kurtosis of -.43 (SE = 0.50).

To test how participants perceived the interviewer's behavior, we regressed perceptions of bias on Behavior (confrontation vs. non-confrontation, dummy-coded), Universalism-Concern (centered) and the Behavior x Universalism-Concern interaction term. As expected, the way participants rated the interviewer's behavior was not predicted by their endorsement of Universalism-Concern or by the behavior of the interviewee. The overall model was not significant (p = .161), and there were no main or interaction effects (p's > .130). Participants generally viewed the interviewer's behavior as biased (M = 6.33, SD = .97, on a 1-7 scale).

[9] We conducted regression analyses using gender and age as control variables. We found no significant effects of either age or gender, and including these variables as controls did not change the pattern of results.
[10] Participants who failed to correctly answer the attention check were evenly distributed across conditions, X^2(1, N = 120) = .10, p = .752.

In order to test our main hypotheses, we regressed behavioral appropriateness on Behavior (confrontation vs. non-confrontation, dummy-coded), Universalism-Concern (centered; M = 4.52, SD = .89) and the Behavior x Universalism-Concern interaction term. The model explained a significant amount of variance, adjusted R^2 = .07, $F(3, 105) = 3.67$, p = .015. There was a significant effect of Behavior on behavioral appropriateness, B = .53, SE = .27, p = .050, η^2_p = .03. The interviewee's behavior was perceived as more appropriate when he confronted (M = 5.02, SD = 1.22) than when he did not confront the interviewer's bias (M = 4.49, SD = 1.57). Importantly, this effect was qualified by the Behavior x Universalism-Concern interaction, B = .81, SE = .31, p = .010, η^2_p = .06 (see Figure 2).

Figure 2. Predicted behavioral appropriateness scores as a function of experimental condition and participants' endorsement of the Universalism-Concern value.



Simple slopes analysis (Aiken & West, 1991) revealed, as predicted, that among participants who scored higher in Universalism-Concern (i.e., one standard deviation above the mean), there was a significant effect of Behavior, B = 1.25, SE = .38, p = .002, η^2_p = .09: participants who scored higher in Universalism-Concern perceived confronting bias as more appropriate (*estimated mean* = 5.45) than not confronting bias (*estimated mean* = 4.21). Among participants who scored lower in Universalism-Concern (i.e., one standard deviation below the mean), however, we found no significant effect of Behavior,

B = -.18, SE = .38, p = .637, η^2_p < .01. As expected, participants lower in Universalism-Concern evaluated non- confrontation (*estimated mean* = 4.90) as appropriate as confrontation (*estimated mean* = 4.73) in these circumstances.[11]

Supplementary analyses revealed that the Behavior x Universalism-Concern interaction remained significant after controlling for the seven other values in the Schwartz scale, B = .81, SE = .32, p = .013, η^2_p = .06. No other scale showed an interaction with Behavior in comparable analyses.

Discussion

Study 2 results supported our hypothesis that individual differences in Universalism-Concern moderate how majority-group observers evaluate responses to bias when the appropriate response is ambiguous. Under circumstances in which the personal costs to a witness for confronting bias are high – in this case, jeopardizing being hired for a job that is needed – participants lower in Universalism-Concern judged non-confrontation as equivalently appropriate as confrontation. By contrast, participants higher in Universalism-Concern, who are highly committed to the value of equality, perceived non-confrontation as less appropriate than confrontation. This effect occurred because participants higher in Universalism-Concern tended to view both confrontation as more appropriate and non-confrontation as less appropriate.

Previous work has demonstrated that greater endorsement of the higher-order value of Universalism predicts a range of behaviors reflecting concern for the welfare of others (Feather & Mckee, 2012; Kuntz, Davidov, Schwartz, & Schmidt, 2015; see also Schwartz, 2010). Although less research has tested the more differentiated values of the Schwartz and colleagues' (2012) revised theory, the evidence that does exist indicates that greater endorsement of Universalism-Concern specifically predicts responses supporting social justice better than other forms of Universalism. In particular, Universalism-Concern is a better predictor of attitudes favoring equal rights for immigrants and homosexuals than Universalism-Nature (which reflects a concern for preserving the natural environment), as well as a better predictor of opposition to

[11] From an alternative perspective, simple slopes analyses demonstrate that participants higher in Universalism-Concern tended to perceive the behavior of the interviewee as more appropriate when he confronted bias, B = .42, SE = .23, p = .068, η^2_p = .03, and as less appropriate when the interviewee did not confront bias, B = -.40, SE = .21, p = .063, η^2_p = .03.

economic inequality than Universalism-Tolerance (which measures a motivation for accepting and understanding people different from the self) (Schwartz et al., 2012; see also Schwartz & Butenko, 2014).

The results of Study 2 offer additional evidence of the distinctive effects of Universalism-Concern for affirming general principles of social justice. Perceptions of how biased the interviewer was did not vary as a function of the participant's endorsement of Universalism-Concern, but differences in this value did affect the way the behavior of the person who witnessed bias was evaluated. Participants who scored low in Universalism-Concern apparently viewed high personal cost for intervention as a mitigating factor for not confronting; they judged non-confrontation to be as socially appropriate as confrontation under these circumstances. The effect remained even when controlling for their endorsement of other values (including Universalism-Nature and Universalism-Tolerance). Thus, in addition to extending work on confrontation of bias by identifying a particularly relevant individual difference variable that moderates evaluations of a decision to confront bias, our findings offer further evidence of the discriminant validity of Universalism-Concern, as distinguished from other forms of Universalism (Nature and Tolerance) and other types of individual values.

Both Benevolence-Caring and Benevolence-Dependability are positively related to Universalism values. However, Benevolence-Caring and Benevolence-Dependability scales were designed by Schwartz et al. (2012) to reflect motivations to care about the welfare of ingroup members and to be a dependable member of the ingroup, respectively. Perhaps because the scenario made social justice concerns more salient than concerns about other ingroup members, we did not find a relation between either of the Benevolence values and the extent a confrontation was perceived to be appropriate.

General discussion

The present studies revealed that the perceived appropriateness of a witness's response to bias depends both on situational and personal factors. In Study 1, majority-group observers evaluated non-confrontation of bias as less appropriate than confrontation when the personal costs of confronting were low but not when the personal costs of confronting were high. These results suggest that when the personal costs of confronting bias are at odds with the social costs of *not* confronting it, the appropriate response to bias is ambiguous.

Building on the results of Study 1, Study 2 investigated how individual differences predict evaluations of responses to bias. In Study 2, observers' personal values associated with equality and social justice predicted their evaluations of confrontations when the appropriate response to bias was ambiguous. In this situation, participants who scored lower on Universalism-Concern evaluate non-confrontation as appropriate as confrontation of bias, while participants who scored higher on Universalism-Concern perceived non-confrontation to be less appropriate.

Taken together, these two studies suggest majority-group observers attend to two different types of costs, and potentially benefits, associated with witnesses' responses to racially biased comments. On the one hand, confronting the biased remark appears to represent a socially valued behavior, particularly by majority-group observers who endorse Universalism-Concern more strongly. Allowing bias to remain unchallenged would permit the violation of basic social principles of fairness and justice, fundamental pillars of society (Oishi, Kesebir, & Diener, 2011; Tyler & Blader, 2003), which would be especially aversive for people highly committed to equality. On the other hand, majority-group observers also appear to attend to the costs a witness of bias could incur for confronting it. Participants, especially those lower in Universalism-Concern, judged non-confrontation to be as socially appropriate as confrontation when the costs for this action were high. For majority-group members less committed to equality and social justice, the costs of confronting bias seem to constitute valid excuses for not confronting it.

To our knowledge, the present research is the first to explore how individual differences shape majority-group observers' perceptions of the appropriateness of different responses to bias, highlighting the important role of endorsing Universalism-Concern. In addition, the current results contribute to the validation of the Schwartz's refined theory of human values by showing that Universalism-Concern, but not other Universalism values, predicts evaluations of behaviors related to social justice. Thus, our studies also underscore the distinctiveness of each of the three factors (Universalism-Concern, Universalism-Tolerance, and Universalism-Nature), which are usually collapsed into a single higher-order value.

We note a seemingly inconsistent finding between Study 1 and Study 2. Across the situations involving potentially high personal costs for confronting bias, we found that participants low in Universalism-Concern in Study 2 viewed non-confrontation slightly

but not significantly more appropriate than confrontation, participants in Study 1 (in which Universalism was not assessed) on average perceived non-confrontation as somewhat less appropriate than confrontation, and participants high in Universalism-Concern in Study 2 evaluated non-confrontation as significantly less appropriate than confrontation. However, this apparent inconsistency may be accounted for by statistical considerations. In particular, while the size of the effect of Behavior (confrontation vs. non-confrontation) on judgments of appropriateness was comparable in Study 2 (η^2_p = .03) and in Study 1 (η^2_p = .02), there were over twice as many participants in Study 2 (*n* = 109) than in the high cost condition of Study 1 (*n* = 45). Statistical power is a function of sample size. Indeed, post hoc estimates of statistical power using G*Power (Faul, et al., 2007) revealed that the power to detect a small-sized effect was .52 in Study 2, but only .25 in Study 1. Thus the non-significant effect for Behavior in in the high-cost condition of Study 1 but significant effect in Study 2 is likely a function primarily of the statistical sensitivity of the test not a difference in the size of the effect.

Taken together, the findings of Study 1 and Study 2 highlight the role of perceived costs and benefits associated with observers' perceptions of majority-group member's decisions about how to respond to bias. Consistent with our hypothesis that individual differences systematically shape how people weigh the costs of confrontation and non-confrontation of bias, Study 2 highlights how endorsement of the value of Universalism-Concern uniquely predicts how people appraise action or inaction by another person who witnesses bias.

Limitations and future research

Methodologically, one limitation of the present research was that although we manipulated the perceived cost to witnesses for confronting bias in Study 1, we did not directly assess participants' perceptions of the costs or benefit to the witness or to society for the alternative behaviors. We did not include such measures before asking how socially appropriate participants perceived the witness behavior because we did not want to sensitize participants to the specific predictions of the work. Measuring perceived costs and benefits after assessing perceptions of social appropriateness may reflect post hoc justifications for ratings of social appropriateness rather than true mediating mechanisms. Moreover, conceptually, observers may not be consciously aware of their processes of weighing different costs and benefits in shaping their assessments (Piliavin et al., 1981;

Nisbett & Wilson, 1977). Nevertheless, future research might consider indirect measures of the attention people devote to either the personal or social costs in their consideration of the situation, for example by measuring the relative cognitive accessibility in a lexical decision task (Kay & Jost, 2003) of words associated with social justice (e.g., equality) and words associated with the consequences for the witness (e.g., employment). These indirect measures would be expected to mediate how people judge the social appropriateness of confronting or not confronting bias.

We also note that we assessed Universalism-Concern, along with other scales in Schwartz and colleagues' (2012) value inventory at the beginning of Study 2, before the manipulation and the assessment of the dependent variables, because it was hypothesized to represent a moderator of the effect of the manipulation of confrontation versus non-confrontation. It is possible that including the value items first in the procedure might operate as a kind of prosocial prime. However, inconsistent with a general prosocial prime interpretation of our findings, Universalism-Concern systematically moderated responses to the manipulation even when controlling for the other values in Schwartz et al.'s instrument. Nevertheless, methodologically, future work might present the value scales at the very beginning and very end of the study (counterbalanced) to assess any order, and potential priming, effects.

Another limitation of the present research is that we relied on a scenario methodology to assess participants' evaluations of a witness' behavior. Indeed, a number of studies of responses to an incident of bias have used scenarios (Ashburn-Nardo et al., 2014; Shelton & Stewart, 2004; Swim & Hyers, 1999; Woodzicka & LaFrance, 2001) or retrospective reports of responses (Good, Moss-Racusin, & Sanchez, 2012; Kaiser & Miller, 2004), in addition to immediate and spontaneous reactions to an incident. We used the scenario methodology for the experimental control it provides for manipulating specific elements of the situation and our interest in a particular outcome – judgments of social appropriateness. Perceptions of what behaviors are socially appropriate or inappropriate are particularly important, because perceptions of normative appropriateness can guide behavior, including intergroup behavior, in ways independent of personal attitudes (e.g., Paluck, 2009).

Nevertheless, because people's descriptions of what they would do in situations portrayed in scenarios do not always match their behavior when presented with the actual incident, particularly in the context of confrontations of bias (Kawakami et al., 2009;

Woodzicka & LaFrance, 2001), future research might productively investigate how people respond to others who do or do not confront bias in more immediately unfolding and compelling situations. In an actual situation directly involving them, observers may spontaneously weigh the social costs of not confronting and the personal costs of confronting differently, in line with the finding that targets of prejudice imagine they would be less influenced by the confrontation costs than they actual are (Shelton & Stewart, 2004). This kind of finding suggests that future work on how observers evaluate whether someone confronts bias should compare the responses of observers detached from the situation, as in the present research, to those more immediate involved in the situation that unfolds.

We acknowledge that the present research focused on responses in a particular social context in which the observer, the person who witnessed bias, and the person who expressed bias were all members of the same group – the majority racial group in the society. Future research might consider how group membership affects observers' evaluations of responses to bias. Both the status of the group and the group membership of the parties may be influential. With respect to group status, majority-group members tend to be less sensitive to prejudice than are minority-group members (Blodorn, O'Brien, & Kordys, 2011; Inman & Baron, 1996) and, as a consequence they may be more likely to view non-confrontation of bias as more socially appropriate.

In addition, whether an observer shares group membership with the target of bias or with the person expressing bias may critically affect the way they evaluate relevant behaviors. Observers who are not members of the group targeted by bias tend to be less supportive of a confrontation than are observers who belong to the target group (Becker & Barreto, 2014; Dodd, Giuliano, Boutell, & Moran, 2001). Also, because people tend to be more lenient with negative behaviors of ingroup members than of outgroup members (Hewstone, 1990), they may be more likely to discount the offensiveness of a statement of an ingroup than an outgroup member, particularly when the comment is directed at a member of a different group (Drury & Kaiser, 2014). To evaluate the potential effects of group membership in this context, future research could parametrically vary whether an observer is a member of the same or a different group as (a) the target of bias, (b) the person who expresses bias, and (c) the witness who does or does not confront bias.

One additional noteworthy finding in the present research, which merits further research, is that across our two studies the ratings of the appropriateness of both

confrontation and non-confrontation were at or above the midpoint of the scale for all conditions. Although confrontation was generally more positively evaluated than non-confrontation on average, participants found any reactions to racism preponderantly appropriate, even if they were high on Universalism-Concern. This result suggests that even those highly motivated by principles of social justice and equality may excuse and see as appropriate other people's inaction (see also Kawakami et al., 2009). Thus, while the finding that confronting bias is perceived as more socially appropriate relative to not confronting bias, the fact that participants assessed non-confrontation as socially appropriate in absolute terms may help account for why even highly principled people often tolerate bias in others (Dickter & Newton, 2013). Whereas past research on this topic has focused primarily on the confrontation of bias, future research might productively consider factors that contribute to perceptions of non-confrontation of bias as socially appropriate. For example, people may generally recognize that confronting bias is personally costly (e.g., Czopp et al., 2006), and this perception of the potential costs incurred may lead them to excuse non-intervention and accept it as socially appropriate.

Conclusion

Our work focused on factors that influence how majority-group observers perceive the social appropriateness of another majority-group member's confrontation and non-confrontation of bias targeting a racial outgroup member. Both situational factors (costs to the witness for confronting bias) and individual differences (in the degree to which observers endorse Universalism-Concern) systematically shape these perceptions. Because people are highly attentive to what is socially appropriate conduct in intergroup behaviors (Blanz, et al., 1997; Gaertner & Insko, 2001; Paluck, 2009; Pereira, Vala, & Leyens, 2009), the way observers' evaluate responses to bias constitutes an important influence that may help or hinder witnesses' confrontation of bias. Considering how observers assess the appropriateness of alternative behaviors helps painting a more complete and complex picture of the processes that may allow traditional biases to persist socially or to combat it through social and interpersonal interventions.

Chapter III

Ingroup favoritism and witnesses' confrontation of racial bias

This chapter is based on:

Lavado, S., Dovidio, J. F., & Pereira, C. R. Ingroup favoritism and witnesses' confrontation of bias. *(Submitted).*[12]

[12] As the first author, I was the primary responsible for the design, implementation, data analysis and writing of the studies reported in this paper. My co-authors (my doctoral advisors) provided fundamental guidance and supervision throughout the entire process, ultimately leading to the submission of this manuscript.

Abstract

The present research, consisting of two experiments, in which group membership was based on assignment to minimal groups, investigated whether the group membership of the person expressing bias against another group (a racial/ethnic group) influences witnesses' confrontation of bias. We predicted that because negative actions tend to be attributed more to internal dispositions of an outgroup than an ingroup member, participants would attribute bias-related characteristics more strongly to an outgroup than an ingroup member who makes a racially biased remark and, consequently, confront him or her more. Moreover, we hypothesized that these effects would occur primarily among participants low in internal motivation to respond without prejudice, and not for those high on this dimension, whose internalized standards transcend group membership. Results converged to support these hypotheses. Practical implications related to the frequency with which people witness bias expressed by an ingroup member, and future directions are considered.

Key-words: confrontation, ingroup favoritism, minimal groups, prejudice, witnesses

Introduction

Much of the literature on confronting expressions of bias has focused on how direct targets (Mallett & Melchiori, 2014; Shelton & Stewart, 2004; Woodzicka & LaFrance, 2001) or members of the group against which bias is expressed (Lee, Soto, Swim, & Bernstein, 2012; Swim & Hyers, 1999; Rattan & Dweck, 2010) respond to a biased statement or action. When these individuals do confront bias, they often elicit negative reactions (Dodd, Giuliano, Boutell & Moran, 2001; Rasinski & Czopp, 2010; Shelton & Stewart, 2004). By contrast, when people who are not associated with the group that is the target of bias confront prejudice, they are less likely to experience backlash and are more likely to be successful at preventing future acts of bias (Drury & Kaiser, 2014; Gulker, Mark & Monteith, 2013; Rasinski & Czopp, 2010). Understanding the factors that may moderate people's willingness to confront bias against a member of another group is thus important, practically as well as theoretically. In the present research, we investigated the role of group membership in witnesses' confrontations of bias. Specifically, we investigated whether the ingroup or outgroup membership of the person expressing bias against another group influences the extent to which people confront bias.

Group membership plays a fundamental role in how people respond to others. In general, people evaluate others whom they perceive to be a member of their group more favorably and are more generous in allocating resources to them, compared to a member of another group (Dovidio & Gaertner, 2010). These effects occur not only in consequential, enduring groups but also in groups in which membership is initially established within a research context and has little consequence outside the laboratory – minimal groups (Tajfel, 1970). Membership in minimal groups can be created, for example, by assigning individuals to groups randomly but ostensibly on the preference for an artist, Klee or Kandinsky (Brewer & Silver, 1978; Tajfel, 1970; Tajfel, Billig, Bundy & Flament, 1971).

One aspect of intergroup relations, which may be particularly relevant to the degree to which people confront an expression of bias, is that individuals systematically make different attributions of negative behaviors performed by ingroup relative to outgroup members. Specifically, people are more likely to attribute the same negative action to the personalities of outgroup than ingroup members, while discounting the culpability of an ingroup member (Hewstone, 1990; Pettigrew, 1979). Because people are usually more willing to confront a negative remark when they perceive it to stem from negative

dispositional qualities, particularly the prejudice of the person who made that comment (Dickter, 2012; Lee et al., 2012), people who witness a biased comment may confront an outgroup member more than an ingroup member.

Although ingroup-outgroup factors may generally affect confrontations of bias, there are also relevant individual differences related to general orientations toward prejudice that could also play a role, potentially in a way that transcends group membership. One such individual difference relates to motivation to respond without prejudice. Plant and Devine (1998, 2009; Plant, Devine & Peruche, 2010) distinguish between internal and external motivation to respond without prejudice. Internal motivation to respond without prejudice, as assessed by their internal motivation scale (IMS), captures the extent people have internalized egalitarian standards, which are used not only to evaluate their own behaviors but also the behaviors of others (Plant & Devine, 1998, 2009; Schmader, Croft, Scarnier, Lickel & Mendes, 2012). External motivation to respond without prejudice, as measured on their external motivation scale (EMS), reflects the extent people are willing to regulate their own behavior to match social norms or to avoid the sanctions of others for being biased. Within the context of our studies regarding responses to other people's bias, we anticipated that internal motivation to respond without prejudice would be primarily relevant. Schmader et al. (2012) found that White participants higher in internal motivation to respond without prejudice reported more negative emotion and showed greater distress-related physiological responses when they witnessed White individuals express racial bias. These responses were unrelated to external motivation to respond without prejudice.

The present research, consisting of two experiments in which group membership was based on assignment to minimal groups, investigated the hypothesis that people would confront a person who expresses racial bias to a lesser degree when that person is an ingroup than an outgroup member and explored (a) the mediating role of attribution of bias-related characteristics (e.g., being close-minded) to the person who made the disparaging remark, and (b) the moderating role of individual differences in internal motivation to respond without prejudice (Plant & Devine, 1998, 2009). Confrontation was conceptualized as the degree of dissatisfaction expressed toward the person responsible for the biased behavior (see Dickter, 2012; Kaiser & Miller, 2004).

We focused on minimal groups to test the psychological impact of group membership per se on this process, independently of variables (e.g., previous negative

attitudes, social status) that may affect confrontation of bias in more enduring groups. However, whereas the group identities of the participant and the person making the biased comment were experimentally defined by assignment to minimal groups, we focused on responses to a form of bias that has important "real-world" consequences – racial bias. Verbal expressions of racial bias perpetuate negative feelings and beliefs against the targeted groups, which have serious consequences for targets' physical and mental health and wellbeing (for recent meta-analysis, see Paradies et al., 2015; Schmitt, Branscombe, Postmes & Garcia, 2014). In a recent study, undergraduates from different ethnic backgrounds reported hearing an average of nine comments targeting other ethnic groups in the course of one week (Dickter & Newton, 2013), revealing that such comments are still prevalent, even in traditionally liberal contexts such as universities.

We hypothesized that group membership (specifically whether a person making a racially biased remark is an ingroup or an outgroup member) and internal motivation to respond without prejudice would jointly influence the degree to which participants would confront the person who made the comment. In particular, we predicted that, because people tend to attribute negative actions more to the character of an outgroup than an ingroup member (Hewstone, 1990; Pettigrew, 1979) and attributions to a person's prejudice are a key factor in confrontation of bias (Dickter, 2012; Lee et al., 2012), participants would (a) generally confront an outgroup more than an ingroup member for making a racially biased remark, and (b) this effect would be mediated by negative dispositional attributions to the commenter, particularly relating to prejudice. However, we anticipated that individual differences in motivation to respond without prejudice would moderate this effect. Specifically, because people who are more internally motivated to respond without prejudice (high IMS) adhere more strongly to anti-bias standards that transcend group membership, they were expected to confront bias to a higher degree, regardless of the group membership of the person expressing the bias.

Study 1

Study 1 investigated how individual differences in internal motivation to respond without prejudice (IMS) and the group membership of the person expressing bias jointly affect the degree to which people confront bias. Participants in Study 1 were led to believe they were interacting with a person who was either an ingroup member or an outgroup member based on the preferences they expressed for the paintings of one of two artists,

Klee or Kandinsky (Tajfel et al., 1971). In the context of a chat-room interaction, participants witnessed the other person making a biased online comment against a racial group to which neither the participant nor the other person belonged (Blacks, or Latinos if the participant identified as Black). We examined participants' chat-room response to that comment and subsequently asked participants to describe their response on dimensions of confrontation and rate the other person's characteristics. These characteristics included general positive and negative qualities, as well as an item, close-minded, which is particularly strongly related to White bias (Sommers & Norton, 2006). Because individuals' perceptions of their own behaviors often differ from those of third-party observers (Pronin, 2009), we investigated our hypotheses with respect to both participants' perceptions of their own responses to the biased comment (i.e., self-reported confrontation) and independent coders' ratings of these responses (i.e., rated confrontation).

We hypothesized that people who are relatively high in internal motivation to respond without prejudice (i.e., high IMS) would confront bias more strongly, because they have strongly internalized egalitarians standards (Plant & Devine, 1998, 2009) and have been shown to respond particularly negatively emotionally when they witness bias (Schmader et al., 2012).

In addition, based on previous research showing that people are more likely to discount negative behaviors and their reflection on the character of ingroup than outgroup members (Hewstone, 1990; Maass Salvi, Arcuri & Semin, 1989; Pettigrew, 1979), we hypothesized that participants would confront an ingroup to a lesser degree than an outgroup member. However, we anticipated that this effect would occur primarily for participants low in IMS, for whom internalized egalitarian standards are relatively weak, and not for those high in IMS, who adhere strongly to egalitarian standards across situations (Plant & Devine, 1998, 2009). In addition, based on previous research supporting the link between perceptions of bias and confrontation (Dickter, 2012; Lee et al., 2012), we expected this pattern of confrontation of bias to be mediated by attribution of bias-related characteristics to the ingroup or outgroup member who made a racially biased remark.

Method

Participants. Participants were 131 adults living in the United States (59 men, 72 women; mean age = 38.94 years, SD = 12.40). The majority of participants (80%) identified as Caucasian/Non-Hispanic White; 6% identified as Black/African-American, 6% as Hispanic/Latino 5% as Asian/Pacific Islander, and less than 1% as American Indian/Alaska Native. 2% of participants choose other identification[13]. Participants were recruited through Amazon's Mechanical Turk. The protocol for the research (for both Study 1 and Study 2) was approved by the Institutional Review Board (IRB) of the university where one of the authors is affiliated.

Procedure. After participants provided their basic demographic information (age, gender, race/ethnicity, education, employment status), we asked them to complete the Kandinsky/Klee preference task, in order to manipulate *group membership*. In the task, participants were presented with a sequence of pairs of paintings from Kandinsky and Klee. For each pair, they were asked to select the painting they preferred. Participants were then assigned either to the Kandinsky or the Klee group randomly, but ostensibly based on their systematic preference for one of the two painters (for a detailed description of the Kandinsky/Klee paradigm, see for example, Bornstein, et al, 1983; see also Tajfel et al., 1971). To reinforce the perceived similarity among group members, we further told participants, "Being a member of the Klees or of the Kandinskys has been shown to relate to personality variables, such as general outlook in life and sociability. In addition, previous research has shown that members of each group tend to find a lot of similarities among themselves and are often surprised to realize they have much more in common than they would have initially thought." In addition, to enhance the impact of the manipulation, we asked participants why they believed they were a Klee or a Kandinsky (depending on their random assignment).

Participants next read that the study was part of a research project in personality examining how people with similar or different characteristics interact together in an online setting (our cover story). They were also informed that they would interact with another person (supposedly also recruited through MTurk) online, who could be a

[13] While it is possible that minority and majority group members react differently to expressions of bias made by a majority group member against a minority group, analyzing those potential differences was not the aim of the current paper, nor did we have enough statistical power to do so. If we include only participants who identified as Caucasian/Non-Hispanic White in the analysis, or only participants who did not identify as Black/African-American, the pattern of results of both Study 1 and 2 remains the same.

member of the same or of a different "artistic preference" group. The "other person" was actually fictitious; the responses made by this person were preprogrammed as part of the experiment. In order to make the interaction more realistic, participants were asked to wait for another person to join them online. After a brief period (approximately one minute), they were told that another person was found.

Participants were then presented the other person's profile. The other person was always portrayed as a 32 years-old Caucasian/Non-Hispanic White, employed for wages and a high-school graduate. We matched the other person's gender to the participant's gender. Embedded in profile was also information about whether the other person belonged to the same or to a different "artistic preference" group (the group membership manipulation). To maintain the salience of group identities, we subsequently referred to the other (fictitious) person as "the other Klee" or "the other Kandinsky" when the person was presented as having the same group membership as the actual participant. If the fictitious person was presented as having a group membership different than that of the actual participant, we used term "the Klee" or "the Kandinsky."

Participants were told that they and the other person would interact together by taking turns answering questions and giving feedback. This was achieved by simulating a chat paradigm in *Qualtrics*, using a sequence of screens and piped text. Participants were always assigned to answer a non-racially relevant question first, and received neutral pre-programmed feedback to their answer, supposedly from the other person. After receiving the bogus feedback, participants were asked to submit a final answer, in which they could choose to incorporate the other person's feedback or not. Then, the roles were reversed, and participants were informed that it would be their turn to give feedback to an answer from the other person in the study.

Next, all participants read that the other (fictitious) person had been assigned the question, "If you could change anything about the way you were raised, what would it be?" The other person's preprogrammed answer represented a biased comment toward a racial or ethnic minority group to which the actual participant did not belong. For Non-Black participants, that group was Blacks. When the participant identified as Black/African-American, the comment was directed toward Latinos. This procedure allowed us to include members of all racial and ethnic groups in our analyses.

The other person's answer to the question about what they would the change about the way they were raised was designed to be disparaging, explicitly prejudiced and

stereotypical. The other person answered, "I was raised in a neighborhood with lots of Blacks [or for Black participants, Hispanics], so I was always scared. I wish I had been raised in a friendlier neighborhood with people who were more like me. Even today I don't like being around Blacks [Hispanics], they are so aggressive all the time."

At this point, the actual participant was asked to give feedback to the other person about his or her answer. To emphasize that how the participant responded could have consequences in this context, the participant was reminded that the other person could change his or her answer after reading the feedback.

After participants submitted their feedback, they were asked to respond to some measures about the interaction (they were assured that their answers would not be shown to the other person). Participants were asked to rate their feedback to the other person. Among filler items intend to distract them from the main focus of the research, participants were asked the extent their feedback was intended to "Be assertive," "Be Critical," "Show displeasure," "Show disagreement," "Be firm," and to "Confront the [[other] Klee/[other]Kandinsky] because you disagreed with what he/she said," from 1 = *not at all* to 7 = *very much* ($\alpha = .87$). In order to maintain the cover story, we also asked participants about the other person's intention when giving feedback.

Participants then rated, from 1 = *not at all* to 7 = *very much*, the extent to which nine different adjectives characterized the other person. Among those adjectives, we included close-minded, the characteristics most associated with White racists (Sommers & Norton, 2006). A principal components factor analysis revealed that these adjectives formed two factors. The first factor, which explained 57.16% of the variance (eigenvalue = 5.72; factor loadings from .74 to .91) comprised six positive adjectives (friendly, funny, outgoing, creative, sensible, fair; $\alpha = .92$). The second factor, which explained 15.22% of the variance (eigenvalue = 1.52; factor loadings .73 and .85) included two negative adjectives that are less related to racism (old-fashioned and strict; $r(129) = .35$, $p < .001$). Close-minded loaded on both factors (loadings of -.52 and .63, respectively for factor 1 and 2); for this reason, we opted to analyze this item separately.

Finally, participants were asked to complete Plant and Devine's (1998) internal motivation to respond without prejudice (IMS) and external motivation to respond without prejudice (EMS) scales (we selected three items of each scale; $\alpha = .91$ and $\alpha = .93$, respectively).

After the study was completed, three independent judges (two female, one male) rated participants' responses to the other person's biased remark in the chat-room. Judges rated the extent, from 1 = *not at all* to 7 = *very much*, to which participants intended to confront the other participant, using two items ("To what extent did the participant intend to confront the other person," "To what extent did the participant intend to disagree with the other person"). The judges' ratings showed good interrater reliability, intraclass correlation coefficient = .86.

We removed from the sample two participants who expressed suspicion about interacting with a real person in their feedback to the other participant. Thus, the final sample consisted of 129 participants[14].

Results

We first analyzed the correlations among IMS, EMS, positive characteristics, negative characteristics, attributions to bias and confrontation of bias (self-reported and rated by judges). The correlation coefficients for these measures are presented in Table 1. As expected, participants higher in IMS (but not participants higher in EMS) reported confronting the biased comment more, and their responses were also rated as being more confrontational. Participants who perceived the other person less positively, more negatively and as more biased also confronted to a higher degree, in both self-reported and rated measures. Consistent with previous work (Devine, Plant, Amodio, Harmon-Jones & Vance, 2002), EMS and IMS were not significantly correlated.

Table 1. Correlations among the measures included in Study 1.

	Confront (self-reported)	Confront (rated)	Close-mindedness	Positive character.	Negative character.	IMS
Confront (rated)	.68**					
Close-mindedness	.55**	.56**				
Positive characteristics	-.36**	-.45**	-.56**			
Negative characteristic	.35**	.33**	.48**	-.20*		
IMS	.22*	.36**	.25**	-.30**	.14	
EMS	.04	-.06	-.10	.14	.02	-.09

Note: * p < .05; ** p < .01

[14] Including these participants in the analysis does not change the results.

About a third of participants reported not confronting the other person (i.e., the mean rating for their answers was of two or less on a scale ranging from 1 = *not at all* to 7 = *very much*). Similarly, judges rated 30% of the responses as not confrontational (corresponding to a score of two or less on a scale ranging from 1 = *not at all* to 7 = *very much*).

Self-reported confrontation. In our minimal group paradigm, half of participants were told that they belonged to the Klees and half participants were told they belonged to the Kandinskys. To ensure that the group label assigned to participants did not influence the results, we first regressed self-reported confrontation, on participant's group label (Klee vs. Kandinsky) and group membership (ingroup vs. outgroup), both dummy coded, IMS (centered; $M = 5.58$, $SD = 1.55$, measured on a 1-7 scale), and their interaction terms. We also included EMS (centered; $M = 3.08$, $SD = 1.95$, measured on a 1-7 scale) and the interaction term of IMS and EMS.

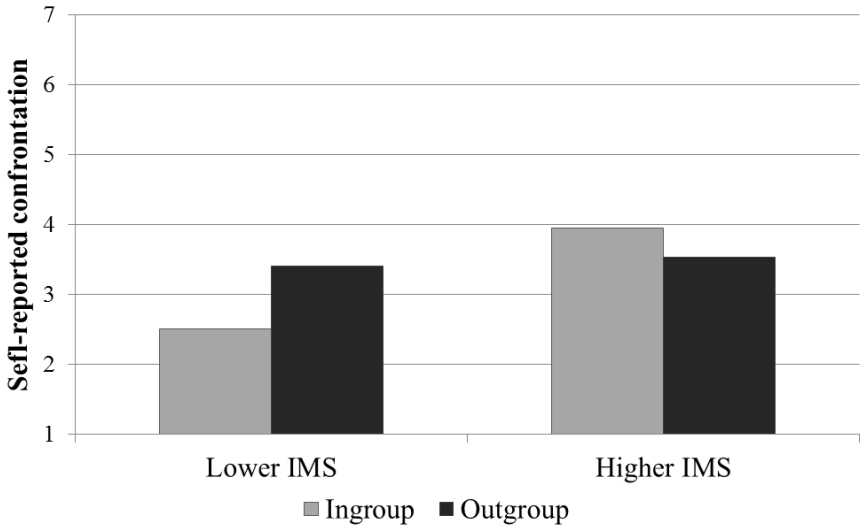
The model explained a significant amount of variance in self-reported confrontation, adjusted $R^2 = .14$, $F(9, 119) = 2.15$, $p = .030$. However, as expected, the group label each participant was assigned (Klee or Kandinsky) did not predict self-reported confrontation ($p = .230$), nor did it significantly interact with the other variables (all $p > .255$). Therefore, we collapsed across participant's group in all subsequent analysis.

Our main prediction was that participants would confront an ingroup member less than an outgroup member. We also expected this effect to be moderated by IMS (but not by EMS). In order to test our predictions, we regressed self-reported confrontation on group membership (0 = ingroup, 1 = outgroup), IMS, and their interaction term. We also included EMS and the interaction between EMS and IMS in the model to control for any effects of EMS.

This model explained a significant amount of variance in self-reported confrontation, adjusted $R^2 = .06$, $F(5, 123) = 2.58$, $p = .029$. Both EMS and the interaction between EMS and IMS were not significant ($p > .400$). There was also no significant main effect of group membership, $B = .26$, $SE = .30$, $p = .382$, $sr^2 = .01$. There was a significant effect of IMS, $B = .44$, $SE = .14$, $p = .002$, $sr^2 = .07$. The higher participants scored on IMS, the more they reported confronting the other person. Importantly, there was a significant interaction between group membership and IMS, $B = -.42$, $SE = .19$, $p = .033$, $\eta^2_p = .03$.

As illustrated in Figure 1, among participants who scored high on IMS (i.e., one standard deviation above the mean), we found no significant effect of group membership, B = -.39, SE = .42, p = .357, sr^2 = .01. Participants who scored high in IMS reported confronting an ingroup member (*estimated mean* = 3.89) as much as an outgroup member (*estimated mean* = 3.50). However, among participants who scored low on IMS (i.e., one standard deviation below the mean) there was a significant effect of Group Membership, B = .93, SE = .43, p = .034, sr^2= .03. Participants low in IMS reported confronting an ingroup member less (*estimated mean* = 2.50) than an outgroup member (*estimated mean* = 3.43).

Figure 1. Interaction between group membership (ingroup or outgroup member) and internal motivation to respond without prejudice (IMS) on predicted scores of self-reported confrontation



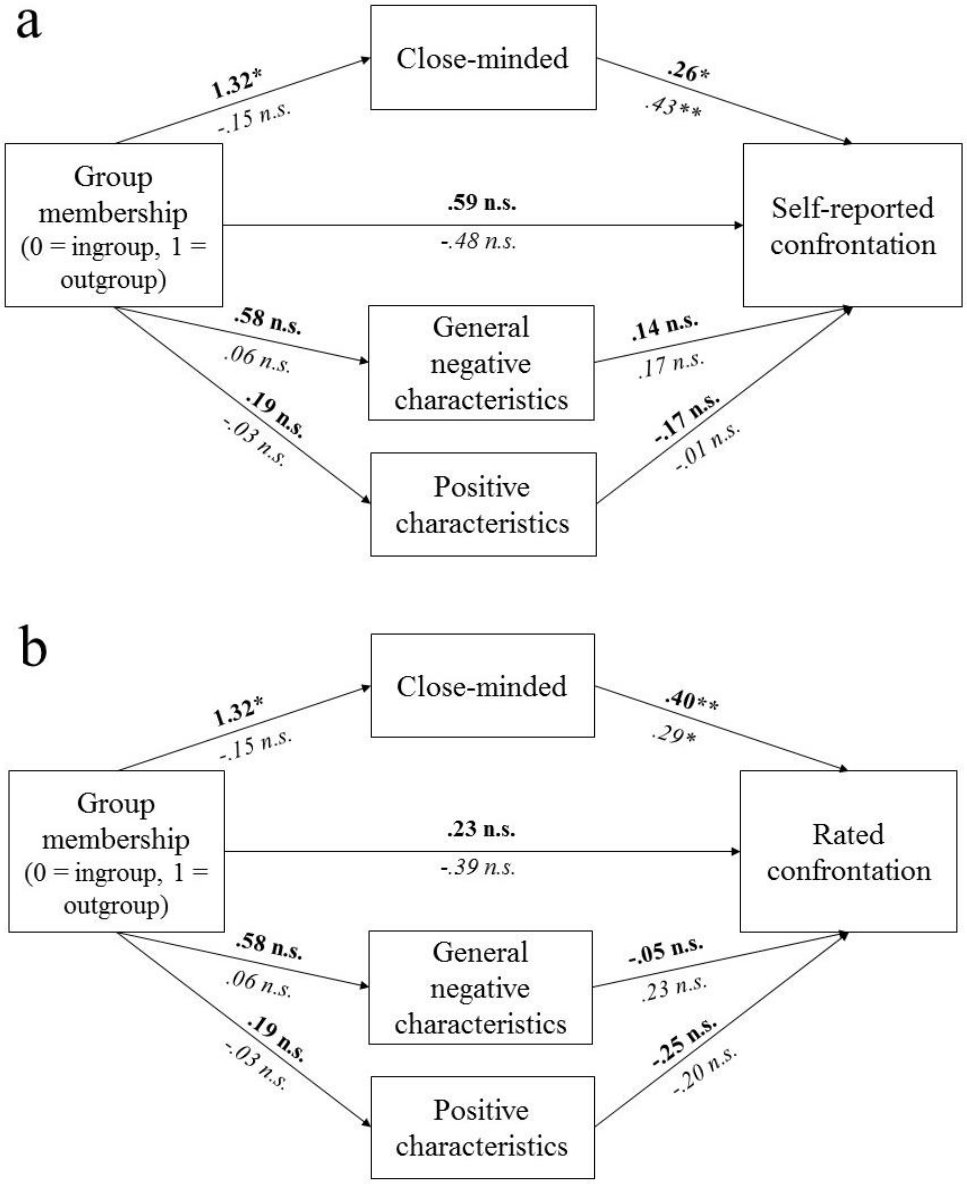
We expected that the extent participants recognized the other participant as biased would mediate the relationship between group membership and self-reported confrontation. Because close-minded is a characteristic more closely associated with White racists than old-fashioned or strict, or any positive characteristics, it more closely captures attributions to biased characteristics. Nevertheless, we also included the general positive and negative scales in the model to test for the effect of close-minded over other general characteristics. Furthermore, we expected the mediation to be significant only for

participants low in IMS, because participants high in IMS tend to be more sensitive to expressions of bias in ways that likely transcend group membership.

We tested this proposed moderated mediation using the PROCESS macro for SPSS (Hayes, 2013). We used Model 8, which allows for multiple mediators and a moderator of the relation between (a) the independent variable and the dependent variable and (b) the independent variable and the mediators. In the model, we included close-mindedness, positive characteristics and general negative characteristics as mediators; we treated IMS as a moderator. We tested the proposed moderated mediation using a bias-corrected 95% Confidence Interval based on 10,000 bootstrap samples.

There was a significant interaction effect between group membership and IMS for close-mindedness, B = -.46, CI = [-.92, -.01]. No similar interaction effects were found for positive characteristics or general negative characteristics. The regression coefficients for participants low and high in IMS are detailed in Figure 2a.

Figure 2. The effect of group membership on (a) self-reported confrontation and (b) rated confrontation is mediated by participants' attribution of a bias-related characteristic (but not by the attribution of general positive and negative characteristics) and this mediation is moderated by IMS.



Note: The bolded values above the line are regression coefficients for participants low in IMS (-1SD); the italicized values below the line are regression coefficient for participants higher in IMS (+1SD). *p < .05, **p < .01, ***p < .001.

As expected, attributions of closemindedness to the other person was a significant mediator of the relation between group membership and self-reported confrontation for participants low in IMS (1 SD below the mean), indirect effect = .43, CI = [.06, .95]. Participants low in IMS attributed more close-mindedness to an outgroup member than to an ingroup member, B = 1.32, p = .011. Attributions of closemindedness further predicted self-reported confrontation, B = .26, p = .036. In addition, there was no evidence that group membership influenced self-reported confrontation independent of its effect on close-mindedness, B = .59, p = .170 among those low in IMS. The same mediation effect was not found for participants high in IMS (1 SD above the mean), indirect effect = -.02, CI = [-.38, .33]. Participants high in IMS perceived an ingroup member as similarly close-minded as an outgroup member, B = -.15, p = .772.

We hypothesized that the attribution of a bias-related characteristic (close-mindedness) would primarily mediate the relationship between group membership and self-reported confrontation. Supporting that hypothesis, attributions of positive characteristics did not mediate the effect of group membership on Self-reported confrontation for participants higher in IMS, indirect effect = .01, CI = [-.09, .15] and lower in IMS, indirect effect = -.02, CI = [-.23, .04]. Likewise, attribution of general negative characteristics did not mediate the effect of group membership on self-reported confrontation, indirect effect = .01, CI = [-.08, .21] and indirect effect = .09, CI = [-.03, .37] for participants high and low in IMS, respectively.

Judges' ratings of participants' confrontation. We expected the results for judges' rating for confrontation to mirror the findings for self-reported confrontation. That is, we predicted that judges would perceive participants' feedback as more confrontational when participants were interacting with an ingroup member than when participants were interacting with an outgroup member. We further hypothesized that this effect would be mediated by participants' perceptions of the other person as close-minded.

We regressed the judges' ratings on group membership, IMS the interaction term of Group Membership and IMS, EMS and the interaction term of IMS and EMS. The model explained a significant amount of variance, adjusted R^2 = .12, $F(5, 123) = 4.37$, p = .001. The pattern of results was similar to the findings for participants' self-ratings. There was no significant main effect of EMS nor an interaction effect of EMS and IMS (both $p > .630$). We also did not find a main effect of group membership, B = .14, SE =

.32, $p = .664$, $sr^2 < .01$. There was a significant effect of IMS, $B = .60$, $SE = .15$, $p < .001$, $sr^2 = .11$. The responses of participants who scored higher on IMS were rated by the independent judges as more confrontational. There was also a marginally significant interaction between group membership and IMS, $B = -.36$, $SE = .20$, $p = .077$, $sr^2 = .02$. Judges' ratings revealed that among participants who scored high in IMS there was no difference between the responses of participants to an ingroup member (*estimated mean* = 4.62) compared to an outgroup member (*estimated mean* = 4.19), $B = -.43$, $SE = .45$, $p = .335$, $sr^2 = .01$. Among participants who scored low in IMS, judges perceived their confrontation of bias expressed by the other person as somewhat, but not significantly, weaker when the other person was an ingroup member (*estimated mean* = 2.72) rather than an outgroup member (*estimated mean* = 3.44), $B = .72$, $SE = .46$, $p = .117$, $sr^2 = .02$.

In the previous analysis, we found that close-mindedness (but not attribution of general positive or negative characteristics) mediated the relation between group membership and self-reported confrontation. We further tested whether close-minded would be a significant mediator of the relation between group membership and judges' ratings of participants' confrontation. In order to do so, we again used the PROCESS macro for SPSS (Hayes, 2013), Model 8, with the same potential mediators and moderator. We used 10000 bootstrap samples and a bias-corrected 95% Confidence Interval.

Results for rated confrontation replicate the findings for self-reported confrontation (see Figure 2b). The attribution of close-mindedness mediated the relation between group membership and rated confrontation for participants low in IMS, indirect effect = .41, CI = [.06, .91], but not for participants high in IMS, indirect effect = -.02, CI = [-.40, .28]. Participants low in IMS perceived an outgroup member as more close-minded than an ingroup member, B = -1.32, $p = .011$; close-mindedness, in turn predicted rated confrontation, B = .40, $p = .003$. Among those low in IMS, there was no evidence that group membership influenced rated confrontation independent of its effect on close-mindedness, B = .23, $p = .602$.

Attributions of positive characteristics and of general negative characteristics did not mediate the effect of group membership on rated confrontation for participants high in IMS (indirect effect = .01, CI = [-.17, .21] and indirect effect = .01, CI = [-.08, .20], respectively) and low in IMS (indirect effect = -.04, CI = [-.29, .06] and indirect effect = .07, CI = [-.05, .36], respectively), suggesting that the attribution of behavior to bias-

related characteristics to the person expressing bias (and not general positive or negative evaluations of the other person) explains why participants low in IMS confronted an ingroup member to a lesser degree than an outgroup member.

Discussion

In the present study, IMS (but not EMS, as expected) moderated the extent people were willing to confront an ingroup (vs. an outgroup) member. Participants who scored high in IMS did not respond differently to an ingroup and an outgroup member who acted biased. Because they have internalized egalitarian standards for behavior, people high in IMS tend to be more responsive to expressions of bias than people low in IMS (Schmader et al., 2012). To the extent they are more sensitive to expressions of bias, people higher in IMS are more likely to attribute those expressions to internal characteristics of the person and to confront more. This finding is also in line with previous research showing that people who are more motivated to promote equality tend to confront bias more (Swim & Hyers, 1999).

By contrast, people low in IMS tended to show ingroup favoritism, confronting an ingroup member less strongly than an outgroup member. Furthermore, perceptions of the other person as close-minded (the characteristic most associated with White racists; Sommers & Norton, 2006) mediated the relation between group membership and both self-reported and rated confrontation for participants lower in IMS. Participants who have weaker internalized egalitarian values were less likely to perceive an ingroup member (vs. an outgroup member) as biased, which, in turn, predicted confrontation of bias. Attributions of positive and negative characteristics less associated with racism were not significant mediators of the relation between group membership and confrontation of bias. These findings further suggest that confrontation is predicted by the extent witnesses make attributions of the biased behavior to dispositional characteristics of the actor, and not by more general measures of liking or disliking.

Even though previous studies have shown that old-fashioned is also a characteristic associated with racial bias (Sommers & Norton, 2006), our results suggest that the old-fashioned and close-minded items measured different constructs. The reason for the difference between Sommers and Norton's (2006) results and our own results may be due to sample differences (undergraduate students and people in public places vs. MTurk workers) or to changes in people's perceptions that occurred in the time interval between

their data collection and ours. Nevertheless, this single-item measure of attribution of bias-related characteristics is a limitation of the current study, which we will address in Study 2.

The degree to which participants confronted was analyzed in two ways: through participants self-reports of the degree to which they had confront and through judges' ratings of participants' actual responses. As mentioned, participants low in IMS reported confronting an outgroup member more than an ingroup member. Although self-reported confrontation and rated confrontation were strongly correlated (see Table 1), the results for rated confrontation showed a similar but non-significant trend (p = .117). The difference between self-reported and rated confrontation results may be due to participants' relying primarily on their feelings and intentions when evaluating their own responses, but judges' relying more on the actual responses gave by participants (Pronin, 2009). Because participants have direct access to their own thoughts and goals, participants' ratings may be more sensitive to contextual factors than judges' ratings. Contextual factors such as the shared membership with the person expressing bias may primarily affect participants' intentions to confront, which are then translated to actual behavior only to a certain extent. Importantly, the effect of group membership through close-mindedness was significant both for self-reported and rated confrontation. As Hayes (2013) highlighted, a direct effect of the independent variable on the dependent variable is not a necessary precondition for mediation analysis: An effect can be transmitted by an independent variable through a mediator on a depend variable without a direct association of the independent and dependent variables.

We note that the attributions of characteristics to the other person were measured after self-reported confrontation. While we argue that attributions of characteristics associated with racism predict confrontation of bias, it is also possible that participants may have made those attributions based on their decision to confront (or not confront) the other person. That is, they may have perceived an ingroup member as less close-minded not because they were less willing to confront him or her. We tested this alternative explanation in Study 2, by asking participants to make attributions for to an ingroup or an outgroup member who expressed bias, without asking them to decide whether to confront or not confront him or her. We also improved the reliability of the measure of attribution of bias-related characteristics to the other person, by including additional adjectives associated with White racial bias (Sommers & Norton, 2006).

Study 2

Study 2 aimed to test the hypothesis that people lower in IMS would attribute an expression of racial bias more to dispositional characteristics of an outgroup member, relative to an ingroup member, even if they did not have the chance to confront that person. Because participants higher in IMS tend to be more sensitive to expressions of bias from another person (Schmader et al., 2012), we expected that participants higher in IMS would generally perceive the commenter as more biased than participants lower in IMS, and that they would not differentiate between ingroup and outgroup members.

In order to test these hypotheses, we again manipulated group membership using a minimal group paradigm. Participants were then presented with a profile, with answers supposedly from a previous participant of the study, who could either be an ingroup or an outgroup member. The actual participants were asked to form an opinion of that person, based on the profile. Among other neutral answers, we embedded in the profile the biased remark used in Study 1. We then measured participants' attribution of bias-related characteristics to the other person, using adjectives generally associated with White racists (Sommers & Norton, 2006).

Method

Participants. Participants were 123 adults living in the United States (57 men, 66 women; mean age = 36.72 years, SD = 12.53). The majority of participants (80%) identified as Caucasian/Non-Hispanic White; 6% identified as Black/African-American, 5% as Asian/Pacific Islander, 6% as Hispanic/Latino and 2% as American Indian/Alaska Native. 2% of participants choose other identification. Participants were recruited through Amazon's Mechanical Turk.

Procedure. After reading that the study was about how people form opinions about others who are similar or different from themselves (our cover story), participants completed the same Kandinsky/Klee preference task we used in Study 1 to establish group membership. After selecting which painting they preferred from a sequence of pairs, we assigned participants (randomly, but ostensibly based on their preferences) to the Klees group or to the Kandinskys groups. Then, they were told that their task was to form an opinion of another person who had participated in a previous survey, based on his or her profile. Participants were presented with the same profile used in Study 1, which provided demographic information for the other person and information indicating whether the

other person was an ingroup or an outgroup member (i.e., a member of the Klee or Kandinsky group). The profile included answers supposedly from the other person, on trivial questions. Among four filler questions and answers, we embedded the same question and biased answer we used in Study 1.

Participants were asked to rate the other person on positive and negative adjectives, using a scale ranging from 1 = *not at all* to 7 = *very much*. We used the same adjectives from Study 1 and added four negative adjectives that were shown to be associated with White racists (insecure, ignorant, naïve, and fearful of change; Sommers & Norton, 2006). A principal components factor analysis revealed a three factor structure. The first factor, which we called bias-related characteristics, explained 38.25% of the variance (eigenvalue = 4.97; factor loadings from .77 to .67) and included close-minded plus the four negative adjectives added in Study 2 (α = .83). The second factor, positive characteristics, explained 14.19% of the variance (eigenvalue = 1.85; factor loadings from .80 to .64) and included the positive adjectives (α = .82). The last factor, general negative characteristics explained 9.22% of the variance (eigenvalue = 1.20; factor loadings from .75 to .59) and included old-fashioned and strict ($r(121) = .36$, $p < .001$). Finally, participants completed the IMS (α = .81) and EMS (α = .93) items used in Study 1.

Results

We began by analyzing the correlations among our variables of interest. As in Study 1, EMS and IMS were not significantly correlated, $r(121) = -.10$, $p = .277$. Participants who scored higher in IMS attributed bias-related characteristics to the other person more strongly, $r(121) = .28$, $p = .002$, but not positive characteristics, $r(121) = -.12$, $p = .183$, or general negative characteristics, $r(121) = .06$, $p = .483$. Conversely, participants who scored higher in EMS attributed more positive characteristics to the other person, $r(121) = -.24$, $p = .007$, but not bias-related characteristics, $r(121) = -.13$, $p = .144$, or general negative characteristics, $r(121) = -.01$, $p = .889$.

We expected, based on the rationale and findings of Study 1, that (a) participants high in IMS would attribute bias-related characteristics to an ingroup member as strongly as to an outgroup member when the person expressed bias; (b) participants low in IMS would, when the other person expressed racial bias, attribute bias-related characteristics less strongly to an ingroup member than to an outgroup member. We hypothesized that

this effect would primarily occur for attributions of bias-related characteristics (compared to attributions of general positive and negative characteristics).

To test the prediction that participants would attribute bias-related characteristics more strongly to the other person when he or she was an outgroup member (vs. an ingroup member), depending on participants' IMS (but not EMS), we regressed negative characteristics on group membership (ingroup vs. outgroup, dummy-coded; 0 = ingroup, 1 = outgroup), IMS (centered; M = 5.77, SD = 1.34, measured on a 1-7 scale), EMS (centered; M = 3.24, SD = 1.85, measured in a 1-7 scale), the interaction term of group membership and IMS, and the interaction term of IMS and EMS. The model explained a significant amount of variance in attribution of bias-related characteristics, adjusted R^2 = .12, $F(5, 117) = 4.38$, p = .001. We found no main or interaction effects of EMS (all p = .330). There was a significant main effect of IMS, B = .64, SE = .17, $p < .001$, sr^2 = .11. The higher participants scored on IMS, the more strongly they attributed characteristics bias-related characteristics to the other person. There was also a significant main effect of group membership, B = .57, SE = .25, p = .017, sr^2 = .04. Participants attributed bias-related characteristics more strongly to an outgroup (M = 4.76) than to an ingroup (M = 4.19). These effects were qualified by a significant interaction between group membership and IMS, B = -.48, SE = .20, p = .018, sr^2 = .04.

As predicted, there was no effect of group membership for participants who scored higher in IMS, B = -.08, SE = .34, p = .827, $sr^2 < .01$. Participants higher in IMS attributed bias-related characteristics to an ingroup member (*estimated mean* = 5.05) as to an outgroup member (*estimated mean* = 4.98) to the same extent. Among participants who scored lower in IMS, we found a significant effect of group membership, B = 1.22, SE = .39, p = .002, sr^2 = .07. Participants low in IMS attributed bias-related characteristics more strongly to an outgroup member (*estimated mean* = 4.55) than to an ingroup member (*estimated mean* = 3.33), replicating Study 1 results.

We conducted a similar regression on positive characteristics. Based on Study 1 results, we did not expect, and did not find, a significant effect of group membership, B = -.18, SE = .21, p = .383, sr^2 = .01, or interaction effect between IMS and Group Membership, B = .24, SE = .17, p = .148, sr^2 = .02. We found a marginally significant effect of IMS, B = -.25, SE = .14, p = .070, sr^2 = .03. Participants higher in IMS attributed less positive characteristics to the other person. We then regressed general negative characteristics (strict and old-fashioned) on the same variables. The model did not explain

a significant amount of variance, adjusted R^2 = -.02, $F(5, 117) = .54$, $p = .745$ and we found no main effects or interaction effect of group membership and IMS (all $p > .300$).

Discussion

The goal of Study 2 was to test differences in the attribution of bias-related characteristics to ingroup and outgroup members and the potential moderating effects of IMS. To maintain the credibility of the study, in Study 1 we measured attributions after participants had the opportunity to respond to the other person in a chat-room paradigm. Study 2 assessed these attributions, the hypothesized mediator of confrontation in Study 1, without permitting an opportunity to confront the racial bias of the other person.

As predicted, IMS (but not EMS) moderated the attribution of bias-related characteristics to an ingroup (vs. outgroup) member, replicating Study 1 results. Participants higher in IMS attributed bias-related characteristics as strongly to an ingroup member as to an outgroup member who openly expressed racial bias. Participants higher in IMS have higher egalitarian standards, and thus they tend to react more strongly to expressions of bias (Plant & Devine, 1998, 2009; Schmader et al., 2012). Because their reaction after witnessing any person expressing bias is stronger, it likely limits ingroup favoritism, especially in a minimal situation such as the one employed in this study.

Participants low in IMS, however, attributed bias-related characteristics more strongly to an outgroup member who expressed bias than to an ingroup member who expressed bias. These results give further support to the hypotheses that participants lower in IMS tend to confront an ingroup member less (as found in Study 1) because they attribute their expression of bias less to his or her dispositional characteristics.

General Discussion

Although societies frequently have egalitarian norms and formal sanctions for exhibiting some forms of intergroup bias (e.g., laws against discrimination toward members of particular groups), individuals are often reluctant to personally confront the bias of others. In addition to feelings of anxiety and even fear, which can inhibit confrontations of bias (Woodzicka & LaFrance, 2001), there are other interpersonal concerns about confronting bias. People who confront bias frequently experience social backlash, under the form of negative emotions and dislike (Czopp, Monteith & Mark, 2006; Dodd et al., 2001; Shelton & Stewart, 2004; Wang, Silverman, Gwinn & Dovidio,

2014). These negative reactions are particularly strong when the person confronting the bias is the target of the negative treatment (Gulker et al., 2013; Kaiser & Miller, 2001, 2003; Rasinski & Czopp, 2010).

Nevertheless, confronting bias is not solely the responsibility of members of targeted groups; members of non-targeted groups may also play an important role in upholding justice norms and principles shared in society and, often, in their group (Drury & Kaiser, 2014). Whereas previous work has focused on differences in reactions to confrontations of bias by people who are or are not members of the target group, the present research investigated whether a shared or different group membership with a person who expresses bias toward another group affects the degree to which witnesses confront bias. We also examined the processes potentially underlying this effect.

Across the two experiments in the present research, we showed that a shared group membership with the person expressing bias and individual differences in motivation to respond without prejudice jointly determined the degree to which witnesses confronted bias. People who are higher in internal motivation to respond without prejudice perceived another person who expressed bias as being more prejudiced, regardless of his or her group membership, and confronted bias to a greater degree. We note that external motivation to respond without prejudice did not have an effect on attributions of biased-related characteristics or self-reported or rated confrontation. This set of findings further supports the fundamentally different motivational processes accompanying internal and external motivation to respond without prejudice, reinforcing the distinctions made in the construction of these measures (Plant & Devine, 1998, 2009). In addition, the present work extends the research of Schmader et al. (2012), which found that people higher in internal (but not external) motivation to respond without prejudice were more emotionally and physiologically distressed when they witnessed racial bias. We demonstrated that differences in emotional and physiological responses between people high and low in internal motivation to respond without prejudice have corresponding attributional and behavioral differences, in terms of attribution of bias-related characteristics to the person expressing bias and stronger confrontations of that person.

Whereas individuals relatively high in internal motivation to respond without prejudice were equally responsive to bias expressed by an ingroup or outgroup member, those relatively low in this motivation responded less negatively – attributing bias-related characteristics less strongly to ingroup than outgroup members who made the same

offensive remark and confronting them less strongly. We note that, in order to control for other factors (e.g., degree of identification, status of the group), we experimentally manipulated group membership using an established minimal group paradigm (Brewer & Silver, 1978; Tajfel et al., 1971). Yet, consistent with other research on attributions for negative or positive behavior (Hewstone, 1990), even such incidental group membership produced weaker dispositional attributions for negative behavior by an ingroup than an outgroup member. Of course, the generalizability of our findings to enduring groups may be moderated by a range of additional factors. For example, people who perceive their group higher in status (Hewstone & Ward, 1985) and those more strongly identified with their group (Johns, Schmader & Lickel, 2005) are more likely to discount the negative behavior of a member of their group in order to maintain their group's positive distinctiveness (Tajfel, 1982). Future research might productively consider how these factors, and others, relate to the way people respond to ingroup and outgroup members who express bias toward another group and explore the extent to which these factors might moderate the responses of even those relatively high in internal motivation to respond without prejudice.

The finding that participants low in IMS may discount biased behaviors from ingroup members (vs. outgroup members), attributing those behaviors less to internal bias-related dispositions of the actor, may contribute to an understanding of why people report relatively low levels of distress after witnessing an ingroup member expressing blatant bias (Kawakami, Dunn, Karmali & Dovidio, 2009). To the extent people attribute the biased behavior of ingroup members to external factors, they would be expected to feel less negative emotions such as shame, an emotion associated with negative events perpetrated by ingroup members (Johns et al., 2005). Future research may investigate whether shame also mediates the relationship between a shared or different group membership with the person expressing bias and witnesses' willingness to confront bias.

We note that the confrontation rate found on Study 1 was somewhat higher than found in previous studies focusing on witnesses' confrontations of bias. While about two third of participants' responses were self-rated or rated by judges as being confrontational at least to some extent, less than one third of witnesses of bias responses' were rated as confrontational in studies using recall or diary methodologies (Dickter, 2012; Dickter & Newton, 2013). The higher confrontation rates found in Study 1 may be due to the interaction paradigm we used. Participants were directly invited to give feedback to the

other person after he or she expressed bias, which constituted a well-defined opportunity to confront seldom found in face-to-face interactions (see Rattan & Dweck, 2010, for the use of a similar paradigm in research on targets' confrontations of bias). In addition, confronting in this situation posed minimal interpersonal costs: the survey was anonymous and participants could not reasonably expect to interact with the other person again. Lastly, the asynchronous nature of the communication allowed ample time for the participant to reflect and prepare his or her response to the biased comment, which may differ from less deliberative responses in more immediate interactive settings (Fazio, 1990).

However, confrontations of bias may occur in a multitude of contexts. Given that social networks such as Twitter or Facebook attract large numbers of users (e.g., Pew Research Center, 2015), people may often face the decision to confront or not confront bias in such computer-mediated settings, which also pose minimal confrontation risk and allow more deliberative responses. Thus, even though generalizations of the current results to confrontations of bias in face-to-face interactions should be done with care, a better understanding of the processes occurring when people witness expressions of bias in computer-mediated settings is relevant in its own right.

In conclusion, the present research showed that a shared group membership with a person expressing bias and internal motivation to respond without prejudice jointly determine witnesses' willingness to confront bias. To the extent people have less internalized egalitarian standards they are more lenient with ingroup (vs. outgroup) members who express bias, perceiving them as less biased and consequently confronting them less. Because people who express racial bias are often friends or acquaintances (Dickter & Newton, 2013) these results importantly contribute to the understanding of witnesses' confrontation of bias, a process that may fundamentally contribute for reducing or maintaining public expressions of prejudice.

Chapter IV

Making us look bad: Group image threat and the normativity of witnesses' confrontations of racial bias

This chapter is based on:

Lavado, S., Pereira, C. R. & Vala, J. Making us look bad: Group image threat and the normativity of witnesses' confrontations of racial bias. *(Submitted).*[15]

[15] As the first author, I was the primary responsible for the design, implementation, data analysis and writing of the studies reported in this paper. My co-authors (my doctoral advisors) provided fundamental guidance and supervision throughout the entire process, ultimately leading to the publication of this manuscript.

Abstract

The present research investigated whether the prescriptive normativity of witnesses' confrontation of bias against immigrants depends on the relationship between the persons expressing bias and the potential confronter. More specifically, we used the self-presentation paradigm to investigate whether it is more normative to confront a close person than a stranger. Study 1 and Study 2 results suggest that confronting is normative both when the person expressing bias is someone close and a stranger, but that confronting a close person is especially normative. Study 3 further shows that the difference in the normativity of confronting a close person and a stranger is explained by threat to the image of a highly-valued group (e.g. groups of friends or family), which legitimizes confrontations of bias. Because prejudice is often expressed by close others, our results contribute to the understanding of normative processes that may reduce expressions of prejudice.

Key-words: confrontation, prejudice, prescriptive norms, self-presentation paradigm, witnesses

Introduction

Despite social norms that condemn expressions of racial bias in many societies (Crandall, Eshleman, & O'Brien, 2002; Dovidio & Gaertner, 2000; Katz & Hass, 1988; Vala & Pereira, 2012), those expressions are still prevalent (Dickter & Newton, 2013; Swim, Hyers, Cohen, Fitzgerald, & Bylsma, 2003), with serious negative consequences for the targets of prejudice (recent meta-analysis, see Paradies et al., 2015; Schmitt, Branscombe, Postmes & Garcia, 2014). One of the ways to reduce future expressions of bias and to motivate amendments from the person responsible for the negative comment is to confront that comment (Czopp, Monteith & Mark, 2006; Mallett & Wagner, 2011), that is, to publicly express dissatisfaction directly to the person being biased (Kaiser & Miller, 2004; Shelton, Richeson, Salvatore, & Hill, 2006). Importantly, confronting bias is not the sole responsibility of people who are, directly or indirectly, targets of bias. It is also the responsibility of majority-group members who witness biased behaviors.

Because non-targets of bias report witnessing several prejudiced comments per week, even in traditionally liberal contexts such as universities (Dickter, 2012; Dickter & Newton, 2013), it is especially important to understand the conditions that facilitate or discourage witnesses' confrontations of bias. Indeed, confrontations of bias can be especially persuasive when they are enacted by witnesses, rather than targets of bias, because it is more surprising and elicits less backlash (Rasinski & Czopp, 2010; Gulker, Mark & Monteith, 2013). One of the factors that may influence witnesses' responses to expressions of bias is the extent the confrontation of such expressions is socially valued and approved, that is, the extent confronting bias is prescriptively normative.

Prescriptive norms and witnesses' confrontations of bias

Prescriptive norms (also called injunctive or subjective norms) reflect the extent behaviors are valued, approved and perceived as desirable by society (Cialdini, Reno, & Kallgren, 1990; Dubois, 2003). These norms have been shown to influence several behaviors, including intergroup behaviors (Blanz, Mummendey, & Otten, 1997; Franco & Maas, 1999; Gaertner & Insko, 2001; Paluck, 2009; Pereira, Vala, & Leyens, 2009). In terms of confrontational behavior, people are expected to be more likely to confront if they perceive such interventions as normative. Accordingly, expectations about how others (including the confronted person) will evaluate the behavior may influence the decision to confront (Good, Moss-Racusin, & Sanchez, 2012; Swim & Hyers, 1999).

Witnesses' confrontations of bias are often positively evaluated, because they address socially unfair behaviors (Dickter, Kittel & Gyurovski, 2012; Lavado, Pereira, Dovidio & Vala, 2016). However, confrontations of bias may arouse anger and defensiveness, which strain social relationships (Czopp, et al., 2006). In addition, confronters are often perceived as impolite, oversensitive or troublemakers (Condor, 2006; Swim & Hyers, 1999; see also Eliezer & Major, 2012; Kaiser & Miller, 2001) and people frequently argue others should "mind their own business" (e.g., Schwartz, 1973). Indeed, while it may generally be socially desirable to confront, several factors may increase or reduce such normativity.

One of the factors that may influence the normativity of confrontational behaviors is the relationship between the person expressing bias and the potential confronter. Indeed, prescriptive norms vary depending on the specific context and the nature of the relationship between the individuals involved. For example, people feel that it is more appropriate to discuss emotionally charged topics with a friend than with a stranger (Clark & Taraban, 1991). Similarly, we predicted that it would be more normative to confront a close person (e.g., a friend, a relative) than a stranger. We further hypothesized that the extent it would be more normative to confront a close person (vs. a stranger) would be explained by two factors: threat to the image of an ingroup and the costs of confronting bias.

Expressions of bias and group image threat. Because expressing bias against outgroups can be considered a failure to meet moral norms regarding fairness of treatment and justice (Dickter, et al, 2012; Effron, Miller, & Monin, 2012; Rutland, Killen & Abrams, 2010) and the moral behavior of group members is central to the formation of group impressions (Leach, Ellemers, & Barreto, 2007; Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012), witnessing an ingroup member expressing bias should threaten the image of the group. One of the possible strategy to deal with threats to the image of an ingroup is to reinforce the ingroup norms (Marques & Paez, 1994; Moreland & Levine, 1982), for example by confronting the person expressing bias.

We hypothesize that threats to the image of an ingroup member would socially legitimize witnesses' decision to confront prejudice. Actions that aim to protect a person's group are generally perceived as socially appropriate and justified (Pereira, et al, 2009; Ramos, Pereira & Vala, 2016), especially when those actions benefit a group that is highly

valued (Kaiser, Hagiwara, Malahy, & Wilkins, 2009). Thus, we predicted that to the extent that expressions of bias threatens the image of a valued group, people would perceive that they have social legitimacy (see Effron & Miller, 2012) to confront.

People tend to value their membership in groups of friends or relatives more than they valued their membership in groups defined by criteria such as nationality (Lickel, et al, 2000). Accordingly, it is possible that only threats to the image of a highly valued group legitimizes, confrontations of bias, increasing the prescriptive normativity of the confrontational behavior. In other words, we expected that only threats to the image of a family or friends' group, and not threats to the image of a social category, such as the Portuguese, would grant social approval to witnesses' confrontation of bias.

The costs of confronting bias. Confronting bias often entail high costs, such as being the target of negative emotions being labeled a "complainer" or even losing a much desired job opportunity (Czopp et al., 2006; Shelton & Stewart, 2004; Swim & Hyers, 1999). The perceived costs of confronting are one of the most important predictors of the people's willingness to confront bias (e.g., Ashburn-Nardo, Blanchar, Petersson, Morris, & Goodwin, 2014; Good, et al., 2012) and, importantly, the costs of confronting bias also predict observers' evaluations of witnesses' responses to bias (Lavado et al, 2016). Even though the costs of confronting appear to be more important predictors of observers' attitudes toward non-confrontations of bias than observers' attitudes toward confrontations of bias, because they make non-confrontations of bias more excusable (Lavado et al., 2016), we explored the influence of this variable on the prescriptive normativity of witnesses' confrontational behavior. We hypothesize that the costs of confronting would reduce the normativity of confronting bias. People tend to anticipate less costs when confronting someone close than a stranger, because reactions of close people tend to be more predictable and less threatening (Ayres, Friedman, Leaper, 2009; Swim, Gervais, Pearson & Stangor, 2009). Thus, the perceived costs of confronting may mediate the influence of the person expressing bias (a close person vs. a stranger) on the normativity of confronting or not confronting bias.

Overview

The present research, consisting of three studies, tested the hypothesis that it would be more prescriptively normative for witnesses' to confront bias when the person making a biased remark is a close person than when the person making that remark is a

stranger. We further predicted that this effect would be explained by differences in the extent people (a) feel that the expression of bias threatens the image of an ingroup, especially the image of a highly-valued ingroup (such as a group of friends or family)and (b) perceive confrontation as costly.

We tested this predictions using the self-presentation paradigm. This paradigm was first proposed by Jellison and Green (1981) and subsequently used on several studies about prescriptive norms. It is based on the finding that people can use their knowledge of social norms to modulate their behavior in order to create positive or negative impressions on others (Jellison & Green, 1981). In this paradigm, research participants are explicitly asked to fill a survey either aiming to gain social approval and/or social disapproval. By doing so, participants should reveal what they think is more normative. A behavior that is more normative should be significantly more endorsed when participants are trying to gain social approval than when participants are trying to gain social disapproval. Several variations of the paradigm have been used, that varied in the design (between or within) and on the potential evaluator of the behavior (a general "other" or a specific person, like a parent or a teacher; for a review of the self-presentation paradigm, see Gilibert & Cambon, 2003; see also Alves & Correia, 2008). In the current paper, we used a between participant design in the first two studies and a within participant design in the third study, maintaining the evaluator unspecified.

In the present studies, we asked participants to imagine hearing a comment that was biased against "immigrants" (Study 1) and "Black immigrants" (Study 2 and 3). We chose these groups because recent reports show that the prevalence of prejudiced behavior (under the form of, for example, verbal harassment in public places) against immigrants in general, but especially against immigrants from African countries, is still relatively high in Portugal, the country where the studies were conducted (Mendes & Candeias, 2013; Santos, Oliveira, Kumar, Rosário, & Brigadeiro, 2006). At the same time, however, there are strong anti-prejudice norms in the country protecting the group (Pereira et al. 2009; Vala, Lopes & Lima, 2008), which may contribute to the normativity of confrontation.

Study1

Study 1 aimed to test the hypothesis that confronting bias is more normative when the person expressing bias is a close person (i.e., a friend, a relative) than when the person expressing bias is a stranger. Participants were invited to take part in a study investigating

"people's opinion about different behaviors in several social contexts." In order to examine beliefs about the normativity of confrontations of bias, participants were asked to imagine several scenarios, where a close person (a friend or a relative) or a stranger (of the same nationality) expressed bias. Participants were randomly assigned to one of three condition. They were either asked to rate how they would act in each scenario in order to present a positive image of themselves (social approval condition); in order to present a negative image of themselves (social disapproval condition); or according to their own opinion (standard condition). We hypothesized that the difference between the social approval and the social disapproval conditions would be larger when the person expressing bias was a close person rather than a stranger, suggesting that it is more normative to confront someone close than a stranger.

Method

Participants and design. 106 undergraduate students completed the survey (53 women, 53 men, M_{age} = 20.72 years, *SD* = 4.39, 93% indicated they had Portuguese nationality, 1% indicated they had more than one nationality and 6% indicated they did not have Portuguese nationality[16]). Participants were recruited through the university, and completed the survey in class with no direct compensation. Participants were randomly assigned to a 3 (instructions: social approval; social disapproval; standard) X 2 (relationship: close person vs. stranger) mixed factorial design where the former is a between factor and the latter is a within factor.

Procedure. Following the self-presentation paradigm, we varied the instruction given to participants in each condition. In the social approval condition, participants were instructed "to answer the survey giving a positive image of yourself, that is, answering in a way that people who read your answers would approve and like you;" in the social disapproval condition, participants were instructed "to fill the survey giving a negative image of yourself, that is, answering in a way that people who read your answers would disapprove and dislike you;" finally, on the standard condition, participants read a standard instruction asking them "to fill the survey according to their own opinion. There are no right or wrong answers, as they all represent your honest opinion."

[16] Excluding participants that indicated a different nationality did not change the results in any of the studies reported in this paper.

Participants were asked to read six small scenarios (presented in appendix C), in which a person made a prejudiced comment against immigrants. Across scenarios, we varied the relationship between the person who expressed bias and the participant. Specifically, in half of the scenarios the prejudiced comment was made by someone close (a relative, a friend or a colleague) and in the other half the prejudiced comment was made by a stranger.

In each scenario, participants were asked to indicate how they would respond to the comment in a 7-point scale ranging from 1= *ignore the comment* to 7 = s*ay the comment was prejudiced*. We submitted the six scenarios to principal axis factoring factor analysis using a promax rotation. The final solution contained two factors accounting for 89.21% of the variance. The first factor included the scenarios where the author of the prejudiced comment was a stranger (loadings ranged from .93 to .96, other items loading < .08). The second factor included the scenarios where the author of the prejudiced comment was a close person (loadings ranged from .87 to .98, other items loadings < .03).

At the end of the survey, participants were asked to provide demographic information (gender, age, nationality) and recall the instructions they received in the beginning of the survey. We removed from the sample 17 participants who failed to correctly recall the instructions[17]. Thus, the final sample included 89 participants.

Results

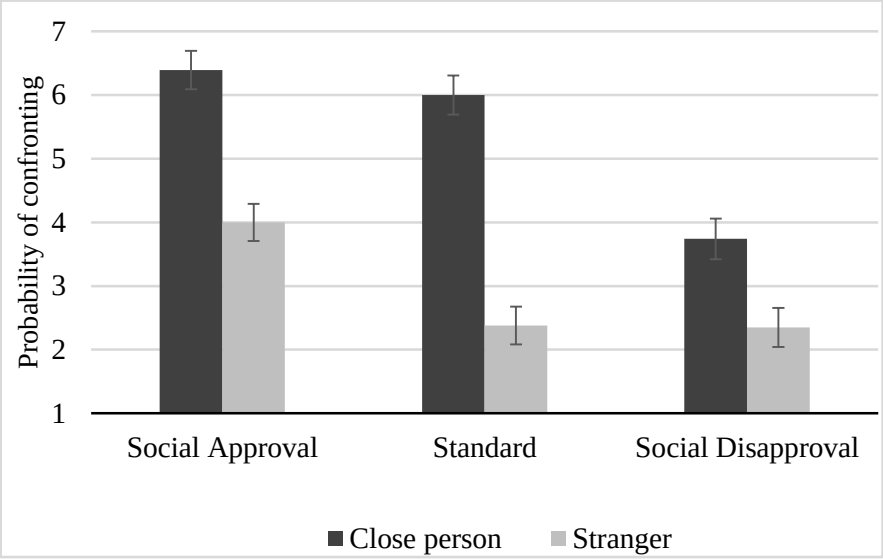
To test the hypothesis that the normativity of prejudice would depend on the relationship between the person expressing bias and the potential confronter, we ran a 3(Instructions: social approval, social disapproval, standard) x 2(Relationship: close person vs. stranger) analysis of variance (ANOVA) with repeated measures on the last variable. The analysis revealed a significant main effect of instructions, $F(2, 86) = 16.01$, $p < .001$; $\eta^2_p = .27$. Participants in the social disapproval condition said they would confront less ($M = 3.04$, $SD = 1.95$) than participants in the social approval ($M = 5.19$, $SD = 1.25$; $p < .001$) and in the standard ($M = 4.19$, $SD = 1.06$; $p = .004$) conditions. These latter two conditions also differed significantly ($p = .009$). We found a main effect of relationship, $F(1, 80) = 219.89$, $p < .001$; $\eta^2_p = .72$. Participants said they would confront

[17] Participants who failed to correctly answer the manipulation check were evenly distributed across instructions conditions, X2(2, N = 106) = 1.75, p = .417.

more when the person expressing bias was a close person (M = 5.42, SD = 2.03) than when the person expressing bias was a stranger (M = 2.93, SD = 1.79).

There was a significant interaction effect between instructions and relationship, $F(2, 86) = 14.74$, $p < .001$; $\eta^2_p = .26$ (see figure 1). When the person expressing bias was a close person, there was a significant effect of instructions, $F(2, 86) = 20.86$, $p < .001$; $\eta^2_p = .33$. Participants in the social disapproval condition said they would confront less (M = 3.74, SD = 2.39) than participants in the social approval (M = 6.39, SD = 1.27; $p < .001$) and in the standard (M = 6.00, SD = 1.19; $p < .001$) conditions, suggesting that confronting a close person is normative. No differences were found between social approval and standard conditions (p = .372).

Figure 1. Means of confrontation for each instruction condition in Study 1.



When the person expressing bias was a stranger there was also a significant effect of instruction, $F(2, 86) = 10.23$, $p < .001$; $\eta^2_p = .19$. Participants in the social approval condition said they would confront more (M = 4.00, SD = 1.69) than participants in the social disapproval condition (M = 2.35, SD = 1.93; $p < .001$) and in the standard (M = 2.38, SD = 1.19; $p < .001$) conditions, suggesting that confronting a stranger is normative. There were no significant differences between social disapproval and standard conditions (p = .940). Thus, the interaction seems to be due to differences in the standard condition. When the person expressing bias was someone close, we found no difference between standard and social approval conditions, while the difference between standard and social

disapproval conditions was significant. A different pattern emerged when the person expressing bias was a stranger: there were no differences between standard and social disapproval conditions, but there were significant differences between standard and social approval conditions.

Importantly, it should be noted that when the person expressing bias was a close person, both social approval and standard conditions means were well above the mid-point of the scale, $t(30) = 10.43$, $p < .001$ and $t(29) = 9.23$, $p < .001$, respectively, while there was no difference between the mean and the scale mid-point in the social disapproval condition, $t(27) = -.58$, $p = .567$. When the person expressing bias was a stranger, there was no difference between social approval condition mean and the scale mid-point, $t(30) < 0.01$, $p = 1.00$, while both the standard and the social disapproval condition means were below the scale mid-point, $t(29) = -7.47$, $p < .001$ and $t(27) = -4.53$, $p < .001$, respectively. These results suggest people are more certain they would and should confront a close person than a stranger.

Discussion

Study 1results offer direct evidence that confronting is generally perceived to be a normative response of witnesses of biased behaviors. Participants instructed to answer in order to obtain social approval said they would confront a prejudiced comment more than participants instructed to try to be socially disapproved, both when the person expressing bias was a stranger and a close person. These results converge with previous studies showing that people evaluate confrontational behaviors more positively than non-confrontational behaviors (Dickter et al., 2012; Lavado et al., 2016). Results further suggest that confronting a close person is more normative than confronting a stranger. When the biased remark was made by a close person, the scores for both the social approval and the standard conditions were well above the scale midpoint, while the scores for social disapproval were at the mid-point. When the person expressing bias was a stranger, the means for all three conditions were below or at the mean point of the scale. These results suggest that people are more certain they should and would confront a friend than they are certain they should or would confront a stranger.

In addition, it is common in the self-presentation paradigm to find no significant differences between the social approval and the standard conditions, because people in the standard condition often try to convey a positive impression of themselves when

answering (Gilibert & Cambon, 2003). Consistent with these assumption, when the biased comment was made by a close person we found no differences in the extent participants giving their own opinion and participants trying to convey a positive image of themselves would confront a biased comment. However, when the person expressing bias was a stranger, there were not only significant differences between the standard and the social approval conditions but also no difference between the standard and the social disapproval condition. Again, these results may suggest that it is more normative to confront a close person than a stranger.

In this first study, we opted for using the general expression "a prejudiced comment" instead of giving examples of actual comments that could be considered biased. We used this strategy because what is considered to be an expression of prejudice may vary from person to person (Sommers & Norton, 2006). By using the expression "a prejudiced comment", we guaranteed all participants imagined they had witnessed an expression of bias directed to immigrants, independently of the specific content of the comment. However, this approach has limitations, such that people may have imagined different comments when the person expressing bias was a close person or a stranger. In addition, in Study 1 we asked people to choose between confronting and ignoring the comment using a single item which may raise concerns about the construct validity of the confrontation measure. Study 2 overcomes these limitations.

Study 2

Study 2 aimed to replicate Study 1 results, again using the self-presentation paradigm (Jellison & Green, 1981). However, Study 2 implemented three important improvements. First, we used a full between factorial design, where both the person who expresses bias (a friend or a stranger) and the self-presentation instructions were manipulated between-participants. Second, we presented participants with only one scenario, where we specified the biased comment made by either a friend or a stranger. Finally, we modified the normativity of confrontation measure in order to in order to allow participants to independently indicate the probability they would verbally confront and abstain from confronting when trying to obtain social approval, obtain social disapproval or convey their own opinion.

Method

Participants and design. The study had a 3 (instructions: social approval; social disapproval; standard) X 2 (relationship: friend vs. stranger) between-factors design. 172 Portuguese undergraduate students completed the study (142 women, 29 men, 1 did not specify gender, M_{age} = 20.30 years, SD = 4.07, 94% indicated they were Portuguese, 3% reported having more than one nationality and 3% indicated they had a different nationality). Participants completed the survey in class with no direct compensation.

Procedure. Participants were randomly assigned to read a scenario where a person, either a friend or a stranger, made a biased comment. The scenario presented to participants was the following (stranger condition in brackets).

> "You are sitting outside a coffee shop with your friends. You are all having a conversation when a black immigrant walks by you. One of your friends [a stranger in the table next to you], who is Portuguese, says "I don't understand how they keep letting immigrants to come and just ruin what is ours."

Participants were asked to indicate how they would react to the situation described in the scenario. Using the same instructions of Study 1, we randomly assigned a third of participants to give their honest opinion about how they would react (standard condition); a third of participants to convey a positive image of themselves (social approval condition); and a third of participants to convey a negative image of themselves (social disapproval condition). Under these instructions, participants were asked to rate the probability they would engage in different behaviors after hearing the biased comment, on a 6-point scale (1- *not very likely*, 6- *very likely*): "say the comment was prejudiced", "show that the comment was incorrect", "ignore the comment", and "change the topic of conversation". The responses were submitted to an exploratory factor analysis (using the principal axis factoring method of extraction), which revealed two factors. The first factor included the items measuring the probability of a verbal confrontation (the first two reported items) and explained 49.70% of the variance (eigenvalue = 1.99; factor loadings .89 and .92). Because they were highly correlated, the responses to these two items were averaged to form a behavior appropriateness scale, $r(170) = .82$, $p < .001$. The second factor included the items measuring the probability of not directly responding to the biased comment and explained 32.75% of the variance (eigenvalue = 1.31; factor loadings .65 and .63). However, this measure will not be further considered, due to the relatively weak correlation between the two items, $r(170) = .39$, $p < .001$. At the end of the survey,

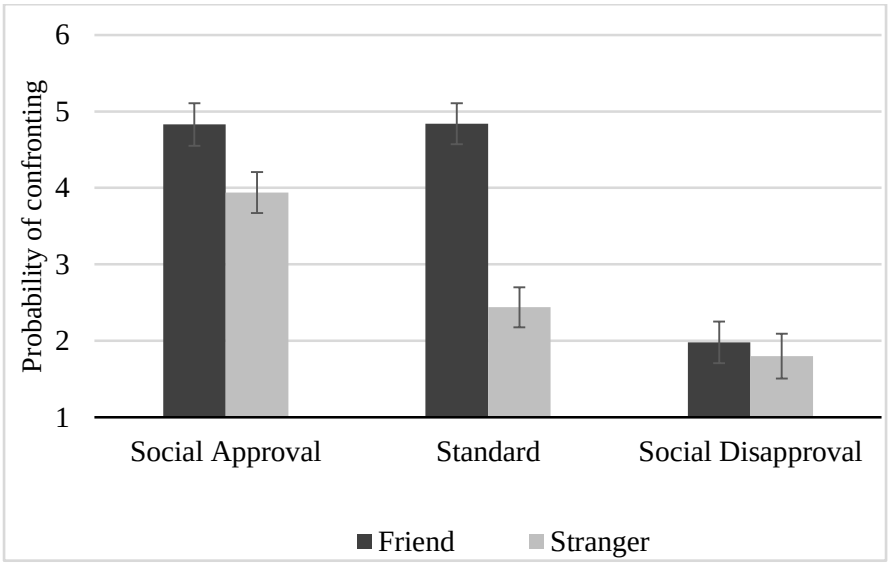
participants were asked some demographic information and to recall the instructions they received in the beginning of the survey. We removed from the sample 22 participants who failed to correctly recall the instructions[18]. Thus, the final sample included 150 participants.

Results

To test our main hypothesis that the normativity of confronting bias would depend on the relationship between the potential confronter and the person expressing bias, we ran a 3(Instructions: social approval, social disapproval, standard) x 2(Relationship: close person vs. stranger) ANOVA. The results, which generally replicated Study 1, are presented in Figure 2. There was a main effect of instructions, $F(1, 144) = 42.64$, $p < .001$; $\eta^2_p = .37$. Post-hoc analysis revealed that participants in the social disapproval condition rated the probability of confronting as lower ($M = 1.89$, $SD = 1.48$) than participants in the social approval ($M = 4.37$, $SD = 1.47$; $p < .001$) and standard conditions ($M = 3.62$, $SD = 1.68$; $p < .001$). The latter two conditions also differed significantly ($p = .017$). In addition, there was a main effect of the relationship, $F(1, 144) = 26.99$, $p < .001$; $\eta^2_p = .16$. Participants indicated they would more likely confront a friend ($M = 3.89$, $SD = 1.84$) than a stranger ($M = 2.77$, $SD = 1.70$).

[18] Participants who failed to correctly answer the manipulation check were evenly distributed across relationship condition, X2(1, N = 172) = .16, p = .690 and instructions conditions, X2(2, N = 172) = 1.27, p = .529.

Figure 2. Means of confrontation for each instruction condition in Study 2.



The analysis also revealed a significant interaction between instructions and relationship, $F(2, 144) = 8.71$, $p < .001$; $\eta^2_p = .11$. When the person expressing bias was a friend, there was a significant effect of instructions, $F(2, 144) = 36.68$, $p < .001$; $\eta^2_p = .34$. Participants instructed to obtain social disapproval said they were less likely to confront a friend ($M = 1.98$, $SD = 1.55$) than participants trying to obtain social approval ($M = 4.83$, $SD = 1.13$; $p < .001$) and answering according to their own opinion ($M = 4.85$, $SD = 1.05$; $p < .001$), and these latter two conditions did not differ significantly ($p = .974$), suggesting that it is normative to confront a friend.

When the person expressing bias was a stranger, there was also a significant effect of instructions, $F(2, 144) = 16.00$, $p < .001$; $\eta^2_p = .18$. When the person expressing bias was a stranger, participants in the social disapproval condition said they were less likely to confront ($M = 1.80$, $SD = 1.42$) than participants in the social approval condition ($M = 3.94$, $SD = 1.64$; $p < .001$). However, participants giving their own opinion said they were as likely to confront ($M = 2.44$, $SD = 1.30$) as participants in the social disapproval condition ($p = .100$) and less so than participants in the social approval condition ($p < .001$).

The obtained results revealed it is normative to confront both a friend and a stranger. However, similarly to the results obtained in Study 1, Study 2 findings suggest that confronting a friend is more normative than confronting a stranger When the person expressing bias was a friend, both the standard and the social approval condition means

on the probability of confronting bias are above the scale mid-point, respectively $t(25) = 6.56$, $p < .001$ and $t(23) = 5.78$, $p < .001$, while the mean of the social disapproval condition is well-below the scale mid-point, $t(24) = -4.90$, $p < .001$. When the person expressing bias was a stranger, however, there was no difference between the social approval condition and the mid-point of the scale, $t(25) = 1.38$, $p = .181$; while both the standard and the social disapproval conditions means are below the scale mid-point, respectively $t(26) = -4.23$, $p < .001$ and $t(21) = -5.63$, $p < .001$, suggesting that participants are more certain they should or would confront a friend compared with a stranger.

Discussion

Study 2 replicated Study 1 results. Participants aiming to obtain social approval said they would confront more than participants trying to obtain social disapproval. This was true both when the person expressing bias was a friend and a stranger, reinforcing the experimental evidence showing it is generally normative to confront expressions of prejudice. However, results also suggest that it is more normative to confront a friend than a stranger. When the person expressing bias was a stranger, the mean for the social approval was at the mid-point of the scale and the mean for the standard conditions was below it, suggesting that participants did not believe very strongly that they would or should confront a stranger. On the contrary, when the person expressing bias was a friend, the mean for both the social approval and the standard condition are well above the mid-point of the scale, suggesting that participants were more certain that they should and would confront a friend who expresses bias.

The results of the first two studies converge to show that confronting expressions of bias from both a close person and a stranger is normative, that is, that confronting is socially valued reaction to behaviors that are not in accordance with egalitarian norms. Results further suggest that it is more normative to confront a close person (vs. a stranger). In Study 3, we explore psychological mechanisms that may account for this result.

Study 3

At least two factors may explain why the prescriptive normativity of witnesses' confrontations of bias depends on who the agent expressing bias is: the perceived costs of confronting and threat to the image of an ingroup – more specifically, a highly-valued ingroup such as a group of friends (Lickel et al., 2000). The costs of confronting were

previously shown to reduce the perceived appropriateness of responses to bias, even though they seem to be more important predictors of attitudes toward non-confrontational behaviors than attitudes toward confrontational behaviors (Lavado et al. 2016). Group image threat is an important predictor of intergroup behavior (Brambilla, Sacchi, Pagliaro and Ellemers, 2013), and because people tend to value actions aiming to protect an ingroup (Pereira et al, 2009), it may legitimize confrontations of bias, especially in highly-valued groups, such as a group of friends (Lickel et al., 2000).

In order to test the hypothesized mediators, we again applied the self-presentation paradigm. However, in the current study, we used a within-participants manipulation (Gilibert & Cambon, 2003), such that all participants answered once trying to gain social approval and once trying to gain social disapproval (the order was counterbalanced). This design allowed us to compute a normativity score for each participant, by subtracting the answers given in the social disapproval condition from the answers given in the social approval condition. The higher the score, the more normative the behavior was perceived to be. We then tested whether, as hypothesized, threats to the image of the group of friends and the costs of confronting mediate the relation between the person expressing bias and the normativity of confrontation of bias.

Method

Participants and design. 82 students completed the study (25 men, 57 women; M_{age} = 22.49 years, SD = 4.06; 96% participants indicated they were Portuguese, 2% indicated they had more than one nationality and 1% indicated a different nationality). Participants were recruited in a university library. All students present at the library at the recruitment time were approached by a female experimenter and asked to complete the survey, which was described as a survey on student's opinions about social situations. They were offered a candy bar in exchange for their participation. About 90% of the approached students accepted participating. The experimenter was unaware of the condition presented in the survey she handled to each participant. The study had a 2(instructions: social approval vs. social disapproval) x 2(order of instructions: negative first vs. positive first) x 2(relationship: friend vs. stranger) mixed factorial design, where the former was a within factor and the latter were between factors.

Procedure. Participants were randomly assigned to read the same scenario used in Study 2 where either a friend or a stranger expressed bias. They were then asked to answer

(from 1 = *disagree a lot* to 6 = *agree a lot*) to four items measuring the extent hearing that person expressing bias would threaten the Portuguese image (two items: Your friend[the stranger]: "threaten the image of the Portuguese people"; "threaten the reputation of the Portuguese people"; $r(80) = .95$, $p < .001$) and the image of their group of friends (two items: Your friend[the stranger]: "threaten the image of your group of friends"; "threaten the reputation of your group of friends"; $r(80) = .88$, $p < .001$), adapted from Brambilla et al.(2013). Participants were also asked to rate the costs of confronting in that situation (3 items: saying that the comment was prejudiced would… 1 = *be very easy.* 6 = *be very hard*; 1 = *be very safe,* 6 = *be very risky*; 1 = *have many positive consequences,* 6 = *have many negative consequences*; $\alpha = .76$).

After, we applied the self-presentation paradigm. Participants were asked to indicate the probability they would confront prejudice if they wanted to convey a positive image of themselves (social approval condition) and if they wanted to convey a negative image of themselves (social disapproval condition). Thus, unlike the previous two studies, participants answered the same scale twice. Half participants read the social approval instructions first and then the social disapproval instructions; and the other half received the instructions in reversed order. Because the self-presentation paradigm posits that normativity is expressed by the difference between the answers given when trying to convey a positive and a negative image of the self, we computed a normativity score, subtracting the answers given in the social disapproval condition from the answers given in the social approval condition. The higher the score, the more normative confronting bias was perceived to be. We excluded four participants whose answers suggested they hadn't understood the instructions (they gave the same, extreme answers both when instructed to answer in order to gain social approval and when instructed to answer in order to gain social disapproval).[19]

Results

We began by analyzing the correlations between the proposed mediators (costs of confronting, threat to the image of the Portuguese and threat to the image of the group of friends) and the normativity of confronting. The correlation coefficients are presented on Table 1. As expected, the lower the perceived costs of confronting and the higher threat

[19] All four participants had been assigned to the friend condition.

to the image of the group of friends, the more normative confronting was perceived to be. There was no relationship between threat to the image of the Portuguese and the normativity of confrontation.

Table 1. Correlation between the proposed mediators and the normativity of confrontation.

	IT: Portuguese	IT: Friends' group	Costs
IT: Friends' group	.38**		
Costs	.04	-.22	
Normativity	-.03	.30**	-.25*

Note: IT = image threat; * = p < .05; ** = p < .01.

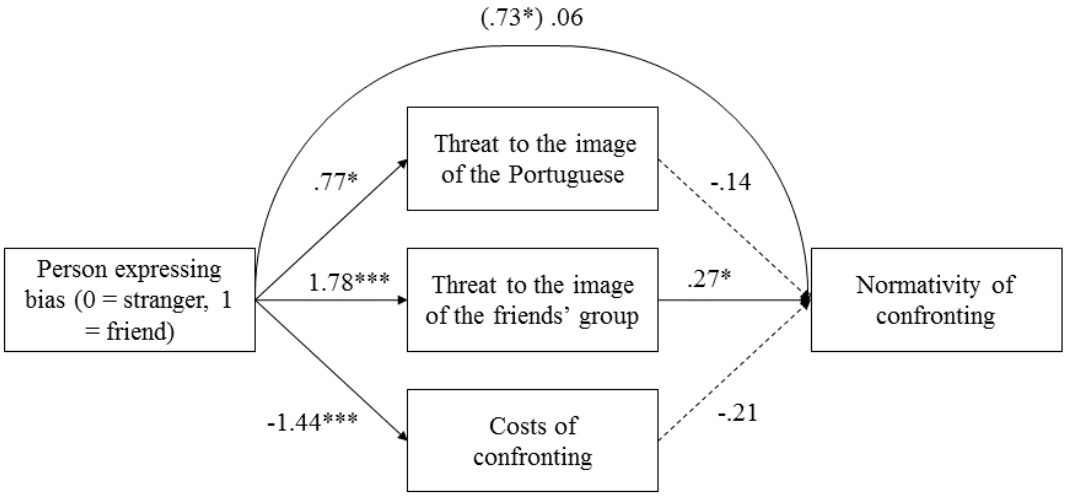
In order to test for effects of the order in which the social approval and social disapproval instructions were presented, we regressed the normativity score on order (dummy-coded, first negative = 0, first positive = 1), relationship (dummy-coded, stranger = 0, friend = 1), and their interaction term. There were no main effect of order and no interaction effect of order with any of the other variables (all $p > .365$). Therefore, we collapsed across groups on this variable for the remaining analysis.

Replicating Study 1 and Study 2, the normativity of confronting was predicted by the person expressing bias, $B = .73$, $SE = .33$, $p = .030$, $\eta^2_p = .06$. The normativity scores were larger when the person expressing bias was a friend ($M = 4.26$, $SD = 1.12$) than a stranger ($M = 3.52$, $SD = 1.72$), suggesting that it is more normative to confront a friend than a stranger.

The relation between the person expressing bias and the proposed mediators was significant. Participants perceived more threat to the image of the of the friends' group when the person expressing bias was a friend ($M = 4.15$, $SD = .24$) than a stranger ($M = 2.37$, $SD = .23$), $B = 1.78$, $SE = .33$, $p < .001$, $\eta^2_p = .27$. Similarly, participants perceived more threat to the image of the Portuguese when the person expressing bias was a friend ($M = 4.31$, $SD = .25$) than a stranger ($M = 3.54$, $SD = .24$), $B = .77$, $SE = .35$, $p = .029$, $\eta^2_p = .06$. Finally, participants perceived less confrontation costs when the person expressing bias was a friend ($M = 2.30$, $SD = .16$) than a stranger ($M = 3.74$, $SD = .16$), $B = -1.44$, $SE = .23$, $p < .001$, $\eta^2_p = .35$.

As detailed in Figure 3, when the normativity score was regressed on relationship, costs, threat to the image of the Portuguese and threat to the image of the group of friends, only threat to the image of the group of friend was a significant mediator of the effect of the relationship between the person expressing bias and the potential confronter on the normativity of confrontation. The mediation effect of threat to the image of the group of friends was reliable with 5,000 bias corrected resamples bootstrapping procedures (indirect effect = .48, 95%CI = .07 to 1.15). Both the mediation effect of threat to the image of the Portuguese (indirect effect = .11, 95%CI = -.44 to .05) and of costs of confronting (indirect effect = .30, 95%CI = -.23 to .97) were not reliable. The overall model explained a significant amount of variance in the normativity of confrontation, adjusted R^2 = .10, $F(4, 73) = 3.07$, p = .021.

Figure 3. The effect of relationship with the person expressing bias on normativity of confronting bias is mediated by group image threat and the costs of confronting.



Note: path values represent unstandardized regression coefficients. The value in the brackets represents the effect of the relationship with the person expressing bias prior to the inclusion of the hypothesized mediators. Solid lines represent significant paths *p < .050, **p <.010, ***p < .001.

Discussion

We hypothesized that the perceived threat to the image of highly-valued ingroup and the costs of confronting would mediate the relationship between the actor expressing

bias against an outgroup and the perceived normativity of confronting that actor. Our results partially supported our hypothesis.

As predicted, we found that a friend who expressed bias threatened the image of the group of friends more than a stranger who expressed bias; and that the higher the threat to the image of that friends' group, the more normative it was to confront. In other words, threat to the image of the friends' group mediated the relationship between the person expressing bias and the normativity of confrontation. These results suggest that a threat to the image of a highly-valued group gives people social legitimacy to act against bias.

The costs of confronting were not significant mediators of the influence of the relationship with the person expressing bias on the normativity of confronting bias. Even though the costs of confronting were perceived to be higher when the person expressing bias is a stranger (vs. a close person) and were negatively related to the normativity of confrontation, they did not significantly explain the relation between who the person expressing bias was and the normativity of confrontation. While previous research has shown that the costs of confronting influence observers' evaluations of different responses to bias, they seem to be more important predictors of observers' attitudes toward non-confrontations of bias than observers' attitudes toward confrontations of bias (Lavado et al., 2016). The current results give further support to those findings.

We also did not find a significant relation between threat to the image of the Portuguese and the normativity of confronting bias. As suggested by previous research, threats to the image of less valued groups such as groups of people of the same nationality (Lickel et al., 2000) may be more commonly addressed by distancing the self from the group (Leach et al., 2007; Lickel, Schmader, Curtis, Scarnier, & Ames, 2005, 2005; Johns, Schmader, & Lickel, 2005) and avoidance of deviant ingroup members (Brambilla, et al, 2013). Therefore, threat to the image of a less relevant group may be a weaker justification for confrontations of bias.

Because in our scenarios both the friend and the stranger were described as Portuguese, we did not expect differences in the extent the two actors would threaten the image of the Portuguese. However, results showed that an expression of bias by a friend threatened more the image of the Portuguese than an expression of bias by a stranger. These results may be explained by the ingroup projection model effect, which suggest that when a superordinate category is made salient (in this case, the Portuguese), people tend to perceive their closer ingroup, as more prototypical of a superordinate category to

which both groups belong (Mummendey & Wenzel, 1999; Waldzus & Mummendey, 2004). It is plausible to think that the more a group member is perceived to be a representative of a group, the more his/hers immoral behavior threatens the group image.

General Discussion

In Study 1 and Study 2, results suggested that although confrontation is generally perceived as prescriptively normative, people perceive a confrontation of a close person as more normative than a confrontation of a stranger. In Study 3 we replicated these results and addressed the mechanism underlying them, showing that confronting a close person is more normative because an expression of bias from a close person (a friend) threatens more the image of a highly-valued group (a friends' group) than an expression of bias from a stranger. Threat to the image of a highly-valued group, in turn, predicted the extent witnesses' confrontation of bias was perceived as a socially approved behavior, over and above other factors such as the costs of confronting and threat to the image of a less valued group. Threats to an ingroup have been shown to legitimize discrimination, by providing seemingly unprejudiced justifications to act on prejudiced attitudes (Pereira, Vala & Costa-Lopes, 2010; Pereira, et al., 2009). On the flip side of the coin, the current studies show that threats to the image of a highly-valued ingroup can also help people justify actions that are consistent with egalitarian norms.

We also explored if the costs of confronting would mediate the effect of the relationship with the person expressing bias on the perceived normativity of confronting. However, the results did not support this hypothesis. Even though participants perceived more costs of confronting when the person expressing bias was a stranger (vs. a close person) and costs were positively correlated with the normativity of confrontation, the effect of costs was non-significant when threat to the image of a highly-valued group was included in the model. However, future research may productively investigate whether the costs of confronting are better predictors of the normativity of non-confrontational behaviors, in line with previous research (Lavado et al., 2016). The present studies contribute to the literature on the normativity of witnesses' responses to bias, by showing that the relationship between the potential confronter and the person expressing bias matters for evaluations of the normativity of confrontational behaviors. Moreover, the finding that threat to the image of an ingroup constitutes an important legitimizing factors of the decision to confront also contributes to our understanding of people's reactions to

being associated with prejudiced behaviors. Previous research has shown that people react strongly when they do not meet their non-prejudiced self-standards (Devine, Monteith, Zuwerink & Elliot, 1991; Monteith, 1993) and when they are accused of racism by others (Czopp, et al., 2006; Shelton, Richeson & Vorauer, 2006). To the extent that people include important others in the self (Aron, Aron, Tudor, Nelson, 1991) and derive part of their self-image from their group memberships (Tajfel & Turner, 1979), protecting the image of those groups (and especially of highly-valued groups) may be a powerful drive of the decision to confront bias. While witnessing a member of an important group may arise feelings of vicarious shame and guilt (Johns, et al., 2005; Lickel et al., 2005), confronting bias may be instrumental in repairing the image of the group and protecting people own image. Although the present research specifically focused on the normativity of confrontations of bias, future studies may productively investigate the influence of perceived image threat on witnesses' actual decision to confront.

Across the three studies, we relied on the self-presentation paradigm (Jellison & Green, 1981) to assess the normativity of confronting. While this paradigm efficiently measures the normativity of different behaviors (Gilibert & Cambon, 2003), it has some limitations. In the case of the present studies, it required participants to image a scenario. In the same way there is sometimes a mismatch between how people believe they would act and their behavior in the actual situation (Kawakami, Dunn, Karmali & Dovidio, 2009; Woodzicka & LaFrance, 2001), there may also be differences between how people imagine others would judge their behavior when reading a scenario and when they are in the actual situation, a question future research might investigate.

In the present paper, we did not investigate whether the degree a comment is perceived to be prejudiced influences the prescriptive normativity of confronting that comment. However, evidence suggest that the more offensiveness a biased comment is, the more positively a confrontation by a witness is evaluated (Dickter et al., 2012). Future research may investigate whether the relation between the person expressing bias and the normativity of confronting is moderated by the extent the comment is subtle or ambiguous. For instance, an ambiguous comment may be perceived as more intentional and less innocuous when it is made by a stranger than when it is made by a close person, increasing the normativity of confronting a stranger. In addition, confrontational behaviors may take several forms, including non-verbal behaviors (Shelton, et al., 2006; Swim & Hyers, 1999). Different confrontational behaviors may be more or less

prescriptively normative (Becker & Barreto, 2014; Dickter et al., 2012). Future research may productively investigate whether the normativity of different confrontational behaviors also depends on the relationship between the confronter and the agent expressing bias.

Conclusion

Confrontational behaviors by witnesses' of bias are prescriptively normative, especially when the person expressing bias is someone close, rather than a stranger. A biased expression from a close person threatens the image of a highly-valued group more than a biased expression from a stranger, and the higher the threat to the image of a highly-valued group, the more confronting bias is normative. Because prescriptive norms are fundamental guides of people's actions toward their own and other groups, influencing behaviors independently from personal attitudes (e.g., Blanz, Mummendey, & Otten, 1997; Franco & Maas, 1999; Gaertner & Insko, 2001; Paluck, 2009; Pereira, et al., 2009), it is important to study the norms regarding witnesses' confrontation of bias, an interpersonal process that may fundamentally contribute to reduce prejudice. Perceptions of threat to the image of a highly-valued group may constitute an important justification for witnesses' decision to confront, granting people social legitimacy to publicly challenge expressions of bias.

Chapter V

Discussion

Overview of the findings

Expressions of racism are still prevalent in societies that have egalitarian norms (e.g., Crandall et al., 2002; Dickter & Newton, 2013; Santos et al., 2005; Vala & Pereira, 2012). Witnesses of those expressions of prejudice are faced with an important decision: to confront the person expressing prejudice, potentially incurring in interpersonal costs (Czopp et al., 2006; Ashburn-Nardo et al, 2014); or to not challenge the behavior, allowing unfair and prejudiced behaviors to persist. Only recently researchers have started to investigate factors that influence people's responses to expressions of bias against members of groups to which the person does not belong (see Dickter, 2012; Dickter & Newton, 2013; Drury & Kaiser, 2014). This is surprising, as witnesses' confrontations of prejudice are more effective in reducing prejudice, because they tend to elicit less backlash and to be more persuasive (Czopp & Monteith, 2003; Gulker et al., 2013; Rasinski & Czopp, 2010). Indeed, witnesses' confrontations of bias may be a fundamental tool to reduce expressions of prejudice against immigrants and, more generally, against racialized groups. Considering that immigration issues are one of the challenges that Europe is currently facing, and will continue to face in the future, it is crucial to investigate strategies that help promoting a more inclusive and egalitarian society (for additional recommendations for promoting social cohesion and positive attitudes toward immigrants, at different geographic levels, see Fonseca & McGarrigle, 2012).

In seven studies, organized into three empirical chapters, we examined the influence of factors related to the relationship between the person expressing racial bias and the witness of those expressions of bias in a) observers' attitudes toward confrontation of bias; b) witnesses' willingness to confront and c) social norms regarding witnesses' confrontations of bias. We further showed that the influence of these social factors in attitudes toward confrontation and on confrontational behaviors is moderated by witnesses' and observers' egalitarian values and standards.

The first empirical chapter (Chapter II) presented two studies that examined observers' attitudes toward the behavior of a witness who, at the end of a job interview, confronts or does not confront a prejudiced remark made by his interviewer. We further manipulated whether the interviewer had low power over the witness (because the interview was for a job the witness did not want or need; low cost condition) or high power over the witness (because the interview was for a much-needed job; high cost condition). Results showed that participants generally evaluated a confrontational

behavior more positively than a non-confrontational behavior, replicating previous findings suggesting that witnesses' who confront are more positively evaluated than witnesses who fail to confront blatantly prejudiced remarks (Dickter et al., 2012). Evaluations of confrontational behaviors were especially positive when the costs of confronting were low, that is, when the interviewer expressing prejudice had low power over the witness. When the costs of confronting were high, a confrontational behavior was only more positively evaluated than a non-confrontational behavior by people who strongly endorsed egalitarian values. People who were less committed to social justice principles did not evaluate a confrontational behavior as more positive than a non-confrontational behavior. We suggested that in the high-cost condition, the appropriateness of responses to a prejudiced comment is more ambiguous, because there is a conflict between the cost to society for not intervening and the cost to an individual for confronting bias. In this ambiguous situation, individual differences in the endorsement of egalitarian values become especially important determinants of attitudes toward different responses to bias. Therefore, in Chapter II, we showed that the endorsement of egalitarian values moderate the extent the power asymmetry of the relation between the potential confronter and the person expressing bias influences observers' attitudes toward responses to racial bias. Only individuals who strongly endorse egalitarian values and are, therefore, more committed to fight bias, believe it is more appropriate to confront even when the costs of doing so are high.

Chapter III described two studies investigating whether witnesses' actual responses to a biased remark revealed ingroup favoritism. More specifically, we hypothesized that witnesses would confront an ingroup member who expressed bias to a lesser degree than an outgroup member who expressed bias. The results supported this hypothesis and further showed that witnesses are less willing to confront an ingroup member than an outgroup member because they attribute less bias-related characteristics to an ingroup member than to an outgroup member who makes the same biased remark. However, only witnesses who were relatively less internally motivated to respond without prejudiced showed this type of ingroup favoritism. People who have internalized stronger egalitarian standards show more distress after witnessing another person expressing biased than people with weaker egalitarian standards (Schmader et al., 2012), in a way that likely transcends group membership. Accordingly, witnesses with stronger motivation to respond without prejudice perceived a biased comment as stemming from internal

dispositions of the person expressing bias as much when that person was an ingroup as an outgroup member.

Finally, the results of the three studies presented in Chapter IV showed that witnesses' confrontation of bias is generally a prescriptively normative behavior, that is, a behavior that is socially valued and approved. However, degree to which confronting is prescriptively normative depends on the relation between the person expressing bias and the potential confronter. Confronting was especially normative when the person expressing bias was someone close rather than a stranger. When a close person (i.e., a friend) expressed bias, it threatened the image of a highly-valued ingroup (i.e., the group of friends) more than when a stranger expressed bias. Threat to the image of a highly-valued group, in turn, predicted the extent witnesses' confrontation of bias is socially approved. Because egalitarian norms suggesting one should confront are often in conflict with social norms suggesting one should "mind their own business," we argued that threat to the image of a highly-valued ingroup provides social legitimacy for witnesses to confront bias.

The findings of Chapter III and Chapter IV may, at first glance, look contradictory. In Chapter III we showed that people tend to confront an ingroup member less than an outgroup member, but in Chapter IV we showed that it is more normative to confront someone close than a stranger. Several factors may account for these findings. We note that in the studies reported in Chapter III, both the ingroup and the outgroup member who expressed bias may be considered strangers: participants' interaction with them before they expressed prejudice was minimal, and participants may not have reasonably expected to interact with them after the experiment was over. Indeed, strangers may often be considered ingroup members at a certain level of inclusiveness (for example, students attending the same university or with the same nationality). Likewise, even though it may be more normative to confront someone close than a stranger when that person is perceived to be biased, it is nonetheless possible that people in the actual situation would discount the behavior of a close person more, attributing biased-related characteristics less strongly to him or her (compared with a stranger). In other words, although it may be more normative to confront someone close *when that person is perceived to be biased*, people may still be less willing to attribute bias-related characteristics to a close person, requiring a more blatantly racist comment to trigger a strong confrontation. This

possibility may help explain the low rates of confrontation of bias found in previous research (Dickter & Newton, 2013), and may be addressed by future research.

While the results of Study 1 of Chapter II suggest that the costs of confronting (emerging from the power of the person expressing bias over the witness of the biased comment) predict observers' evaluations of witnesses' responses to expressions of bias, in Study 3 of Chapter IV the costs of confronting were not significant predictors of the normativity of confronting a close person (vs. a stranger), when threat to the image of a highly-valued ingroup was included in the model. However, we note that in Chapter II the costs of confronting increased the perceived appropriateness of not confronting bias, but did not significantly decrease the perceived appropriateness of a confrontational behavior. That is, the costs of confronting were a more significant predictor of observers' attitudes toward non-confrontations of bias than of observers' attitudes toward confrontations of bias. This finding may help explain why the costs of confronting were not significant predictors of the extent a confrontation of bias was perceived to be socially valued and approved, as the results of Study 3 of Chapter IV suggest.

Confrontations of prejudice help reduce expressions of prejudice against immigrant groups. Public expressions of prejudice against immigrants perpetuate negative attitudes and feelings toward the group (Blanchard et al., 1991; 1994), which, in turn, increase majority-group members' support of restrictive laws and policies regarding immigration (Lima-Nunes, Pereira & Correia, 2013; Pereira et al., 2010). Challenges to those public expressions may lead to more favorable attitudes toward immigrants (Rasinski & Czopp, 2010), promoting a more welcoming and integrative society. This thesis, developed in the context of a PhD program in Migrations, aimed to contribute to our knowledge of confrontations of bias, by focusing in confrontations of expressions of prejudice directed to immigrants and also including expressions of prejudice against black people who may or may not have citizenship status. In Portugal, people often assume that anyone who lives here but is not White is "a foreigner", regardless of the person's place of origin or citizenship status (that is, there is a racialization process where group boundaries are defined by phenotypical characteristics; Cabecinhas, 2002). It is also likely that public expressions of prejudice would target people more due to their skin color than due to their citizenship status (see the report of the European Union Agency for Fundamental Rights, 2009). In the US, the concept of race is ingrained in everyday practices and discourse (e.g., Dovidio & Gaertner, 2010; Essed, 1991). The citizenship status of the person

targeted by prejudice may nuance the expressions of bias and the confrontational (or non-confrontational) responses; however, it would not have been as productive to attempt to capture those nuances in this early phase of the research.

In the present thesis, we aimed to investigate the social and individual factors that facilitate or hinder witnesses' confrontations of bias. The results of our empirical studies converge to suggest that the relationship between the person expressing bias and the person witnessing that expression influence witnesses' confrontation of bias at the attitudinal, behavioral and normative levels. However, majority-group witnesses' and observers' egalitarian values moderate the influence of those social factors. Non-targets of prejudice with strong egalitarian motivations are more supportive of witnesses' confrontation of bias and are more willing to confront, regardless of who the person expressing bias is.

Theoretical and practical contributions

The majority of studies investigating confrontations of bias has focused on the responses of people who are either directly or indirectly (due to their group membership) targeted by the biased expression (e.g., Becker & Barreto, 2014; Dodd et al., 2001; Good et al., 2012; Swim & Hyers, 1999), with few studies directly investigating witnesses' confrontations of biased remarks. However, the responsibility of showing dissatisfaction with public expressions of prejudice, promoting egalitarian norms should be shared by both targets and witnesses of those expressions. In the current thesis, we helped to fill that gap in the literature, investigating the joint effect of social and individual factors that influence witnesses' responses to bias, and attitudes and prescriptive norms regarding those responses.

The results obtained in the empirical studies that comprise the current thesis suggest that the relationship between the person expressing bias and the potential confronter is a fundamental factor in witnesses' confrontation of bias. We showed that different facets of the relationship between the person expressing bias and the witness of that expression – namely, the power of the person expressing bias over the confronter, a shared group membership and the degree of proximity between the confronter and the person expressing bias – not only influences witnesses' actual responses to bias, but also observers attitudes and social norms regarding those responses. While previous research had already showed that power and proximity are important factors in confrontations of

bias (Ashburn-Nardo et al., 2014; Ayres et al., 2009; Shelton & Stewart, 2004), we further showed that those factors influence, respectively, attitudes and norms regarding witnesses' confrontation of bias. In addition, to our knowledge, no previous research had investigated the influence of a shared group membership with the person expressing bias on people's willingness to confront. The group membership of the person expressing bias is a particularly important factor in witnesses' confrontation of bias, considering that often[20] both the person expressing bias and the witness of that prejudice remark are members of high-status groups (Dickter & Newton, 2013; Inman & Baron, 1996; Swim, Hyers, Cohen, & Ferguson, 2001).

Practically, our finding that the attribution of biased-related characteristics to the person expressing bias predicts witnesses' willingness to confront the biased remark (see Chapter III) corroborates the argument that, in order to increase witnesses' confrontation of bias, it is important to encourage them to recognize prejudiced remarks (Drury & Kaiser, 2014). People who do not belong to the targeted group tend to acknowledge prejudice less than targets of that prejudice, especially if the expression of bias is subtle or ambiguous (Drury & Kaiser, 2004; Salvatore & Shelton, 2007). Previous research has already highlighted that promoting awareness of people's own bias may be an important tool to reduce prejudice (Perry, Murphy & Dovidio, 2015); our results complement these findings by suggesting that being aware of other people's bias may also be important to reduce expressions of prejudice. However, confrontations of bias may be especially effective if they, rather than generating high defensiveness on the part of the confronted person, promote reparation efforts (Mallet & Wagner, 2011). Therefore, it may be important not only to help people to recognize their own biases (Perry et al., 2015) but also to confront more constructively (see Crosby, 1993), perhaps aiming more to change prejudiced beliefs and less to accuse someone of being racially biased (which is highly threatening; Van Dijk, 1992; Winslow, 2004).

In Chapter II and Chapter III of the current thesis, results showed that the extent people endorse or have internalized egalitarian values moderates the effects of social factors on observers' attitudes toward responses to bias and witnesses' actual responses to bias. Participants who strongly endorse universalism-concern (a value related to social

[20] We of course acknowledge that the person making a prejudice remark may also belong to another low-status group, or even to the group targeted by prejudice (e.g., women who stereotype other women as being less competent than men or who assume that a person in a high power position is male); and that the witness of the biased comment may also belong to a (non-targeted) low-status group.

justice and equality) supported a witness who confronts bias even when the costs of confronting were high; and participants who were strongly internally motivated to respond without prejudice confronted an ingroup member as much as an outgroup member. While previous research have investigated the role of attitudes toward the targeted group on confrontational behavior (Ashburn-Nardo et al., 2014; Dickter, 2012; Dickter & Newton, 2013), we are not aware of any research that directly investigated the role of egalitarian values confrontations of bias, even though values are conceptualized as being less situation-specific that attitudes (Schwartz, 1992) and, therefore likely comprise the specific attitudes studied in previous research.

Besides demonstrating that people with stronger egalitarian values are more resilient to social factors that influence responses to bias and attitudes toward those responses, we also directly contributed to the literature on Schwartz's theory of basic human values (Schwartz, 1992; Schwartz et al., 2012) and on motivations to respond without prejudice (Plant & Devine, 1998; 2009). Schwartz et al (2012) have recently proposed a refined version of the theory of basic human values, in which they distinguish between three lower order universalism values: universalism-concern, universalism-tolerance and universalism-nature. We further contribute for the validation of the theory by showing that only universalism-concern, and not universalism-tolerance or universalism-nature, predicts observers' attitudes toward witnesses' responses to bias (see Chapter II). Similarly, our results showing that internal motivation to respond without prejudice is predicts the attribution of bias-related characteristics to the person expressing bias and, consequently, to witnesses' willingness to confront bias (see Chapter III) contribute to a recent line of research showing that people who are internally motivated to be egalitarian are, besides being more motivated to control their own prejudice (Devine et al., 2002; Plant & Devine, 2009), more responsive toward other people's expressions of prejudice (Schmader et al., 2012). Practically, these results suggest that, in order to increase the prevalence of witnesses' confrontations of bias, it is important to promote the endorsement and internalization of egalitarian beliefs and standards.

Fear of negative backlash, from the confronted person but also from observers, may be one of the reasons why witnesses' refrain from confronting. While this fear might be sometimes justified (Dodd et al., 2001; Swim & Hyers, 1999), our results suggest that the prescriptive norms and the attitudes toward witnesses' confrontations of bias are generally positive (see Chapter II and Chapter IV of the current thesis; see also Dickter

et al., 2012). That is, confronting bias may be interpersonally less costly than it is anticipated (Swim & Hyers, 1999), at least to the extent observers are concerned (if not to the extent the confronted person is concerned; Mallet & Wagner, 2011). As mentioned previously, promoting a higher awareness of self-biases (Perry et al., 2015) might also reduce the extent the person expressing bias reacts negatively to being confronted about a specific remark, therefore reducing the perceived costs of confronting bias and, consequently, increasing confrontations of bias (Ashburn-Nardo et al., 2014; Good et al., 2012; Shelton & Stewart, 2004).

In societies where, even after decades of egalitarian norms, biases against people with different origins and ethnicities still persist, it is increasingly important that people not only internalize unprejudiced self-standards but also take an active role against prejudice, collectively or individually. Witnesses of bias can play a fundamental role in reducing biased expressions, by effectively challenging them. Indeed, open discussion and acknowledgement of bias might pave the way for the construction of a more egalitarian society (Monteith, 1993; Perry et al., 2015). Thus, our results inform theories and practices regarding witnesses' confrontations of bias, a process that may fundamentally reduce expressions of prejudice against racialized groups such as immigrant people, promoting a more fair and inclusive society.

However, we also wish to corroborate Drury and Kaiser's (2014) recommendation that witnesses of bias should work against prejudice alongside with targets of bias, instead of on their behalf. While witnesses' may be important allies in addressing expressions of prejudice, the kind of paternalistic helping that reduces targets' ability to deal with the prejudiced remark in their own terms, no matter how well-intentioned, may do more harm than good. That is, witnesses' of prejudice should use their voice aiming to reduce prejudice and supporting others who do so, without speaking over the voice of members of the target group.

Limitations and future directions

One of the novelties of the current thesis, the focus on witnesses' responses to bias and on observers' attitudes toward those responses, may be considered one of its limitations. We do not know if the variables we investigated would similarly influence the attitudes and behaviors of members of target groups. Minority-group members tend to be more responsive to expressions of bias than majority group members (Drury &

Kaiser, 2014; Inman & Baron, 1996) and people who belong to the target group tend to be more supportive of a fellow ingroup member who confronts than people who do not belong to the target group (Becker & Barreto, 2014; Dodd et al., 2001). These differences suggest that observers' who belong to the target group might likewise be more supportive of witnesses' confrontations of bias, being less influenced by contextual variables. However, future research is necessary to investigate whether this is the case. Similarly, future research may investigate whether members of the target group also display ingroup favoritism when a shared group membership with the person expressing bias is salient, or if, on the contrary, they tend to confront more an ingroup member than an outgroup member, perhaps because they anticipate less backlash.

Other variables, such as ingroup identification, that have been shown to be associated with targets' confrontation of bias may have a different influence on witnesses' confrontation of bias. Targets of bias tend to confront more when they more strongly identify with the group (Ayres et al., 2009; Good et al., 2012), and targets more identified with the group also tend to evaluate an ingroup member who confronts more positively (Becker & Barreto, 2014; Kaiser et al, 2009). However, the effect of group identification for majority-group members who witness another majority-group member expressing bias may depend on the perceived severity of the biased comment. Johns et al (2005) showed that highly identified majority-group members (relatively to less identified majority-group members) responded more negatively to an ingroup member who acted biasedly only when the event was very negative. However, when the event was less negative, participants more identified with the group showed less negative reactions than participants less identified with the group (Johns et al., 2005; these results are also consistent with research on the subjective group dynamics theory and the black sheep effect, Branscombe, Wann, Noel, & Coleman, 1993; Marques & Paez, 1994; Pinto, Marques & Paez, 2015). For confrontations of bias, this may mean that more identified witnesses are less willing to confront an ingroup member who expresses more subtle or ambiguous bias, but are more willing to confront an ingroup member who expresses blatant bias.

Relatedly, in the studies that comprise this thesis, we always either asked participants to imagine they had heard "a prejudice remark against immigrants" (Study 1 of Chapter IV) or presented participants' comments that were designed to be blatantly

prejudiced and stereotypical of ethnic minorities or immigrants[21]. Ambiguity of the biased comment has been shown to influence both attitudes toward responses to that comment (Dickter et al., 2012; Dodd et al., 2001) and actual confrontations of the biased remark (Dickter, 2012; Dickter & Newton, 2013; Lee et al., 2012). People might attribute bias-related characteristics less strongly to a person who expresses ambiguous rather than blatant prejudice and, consequently, be less willing to confront him or her (see Chapter III; Lee et al., 2012). Accordingly, because individual differences are stronger determinants of behavior in more ambiguous situations (Mischel, 1973; Snyder & Ickes, 1985) it is expectable that the endorsement and internalization of egalitarian values would be more relevant in witnesses' confrontation of subtle biases. Future research may investigate whether the ambiguity of the biased remark moderates the effect of the individual and social factors that were the investigated in the current thesis.

Across the studies presented in this thesis, we mainly focused on witnesses' confrontations of bias against Black people. The choice of Blacks as the group targeted by prejudice has important theoretical and practical implications, considering that Black people are simultaneously one of the groups more protected by anti-discrimination laws and social norms (Crandall et al, 2002; Vala & Pereira, 2012) and one of the main targets of prejudice both in Portugal and in the US (Santos et al., 2005; Dickter & Newton, 2013). Even though we would expect our findings to generalize to other low-status groups who are equally protected by social norms, different variables and processes may account for attitudes, norms and behaviors in responses to expressions of prejudice targeting groups against which is more socially accepted to discriminate (see Crandall et al., 2002).

In the studies presented in Chapter II and Chapter IV we relied in scenarios to investigate individual attitudes and social norms regarding witnesses' confrontations of prejudice. This methodology has been used in several studies examining attitudes toward people who confront (or not confront) bias (Ashburn-Nardo et al., 2014; Becker & Barreto, 2014; Kahn, Barreto, Kaiser & Rego, 2015), because it provides experimental

[21] In Study 1 and Study 2 of Chapter II the biased remark was "I really liked you and I think you and the company would be a good fit. We had a lot of Black applicants, so it's good to have someone White for a change. I'll contact you when we have a decision." In study 1 and Study 2 of Chapter III, it was "I was raised in a neighborhood with lots of Blacks [or for Black participants, Hispanics], so I was always scared. I wish I had been raised in a friendlier neighborhood with people who were more like me. Even today I don't like being around Blacks [Hispanics], they are so aggressive all the time;" and finally, in Study 2 and Study 3 of Chapter IV, it was "I don't understand how they keep letting immigrants to come and just ruin what is ours."

control to manipulate the specific aspects of the situation that are the main focus of the study. However, how people imagine they would behave in a situation described in a scenario often differs from their reaction in the actual situation, particularly in the context of responses to bias (Kawakami et al., 2009; Shelton & Stewart, 2004; Woodzicka & LaFrance, 2001). Similarly, the extent to which people imagine they would be have favorable attitudes toward another person's response to bias may be different from their attitudes in a more compelling situation. The weight given to factors such as the power of the person expressing bias or the degree of proximity between the person expressing bias and the confronter may be different for people more detached from the situation, compared with observers in the more immediate situation. In addition, majority-group members report that they will be more emotionally upset by witnessing an act of bias than they experience in the actual situation (Kawakami et al., 2009). If people closer to the immediate situation experience less upset in response to a racially biased comment, they would also likely evaluate a confrontation by a witness less positively. Thus, future research on how people judge the appropriateness of witnesses' decision to confront or not confront bias might consider how the proximity to the actual incident and the associated salience of different factors may moderate attitudes and norms regarding witnesses' responses to bias.

We likewise note a methodologic limitation of Study 1 of Chapter III, in which we investigated witnesses' actual responses to prejudice using a chat-room paradigm. Interactions in chat-rooms such as the one used in that study have important characteristics that distinguish them from face-to-face interactions. First, in our study, the interaction was anonymous; participants were not ask to convey any information about themselves other than basic demographic information. Second, participants could not reasonably expect to interact with the (fictitious) person expressing bias after the experiment was over. Third, the asynchronous nature of the interaction in such computer-mediated settings allowed participants to take time to reflect and prepare their answers to the biased comment. Lastly, participants' were directly invited to respond to the other person after he or she expressed bias. These characteristics are rarely (if ever) present in face-to-face interactions. Studying responses to bias in computer mediated settings is important – increasingly large number of people use social networks such as *Facebook* or *Twitter* (Pew Research Center, 2015) and it is likely than people often face the decision to confront or not confront a biased comment in such settings. However, the processes

involved in responses to bias are likely moderated by contextual variables. While we would hypothesize that witnesses' willingness to confront in a face-to-face interaction is also influenced by a shared group membership with the person expressing bias (as we showed in Chapter III), especially when that membership is highly salient, future research should test this assumption.

There are reasons to believe that prescriptive norms regarding witnesses' responses to prejudice influence witnesses' willingness to confront when faced with expressions of bias. Prescriptive norms have been shown to influence a variety of behaviors, such as environmental behaviors (Cialdini, et al., 1990; Kallgren, Reno & Cialdini, 2000), prosocial behaviors (e.g., Gruder, Romer & Korth, 1978; Rutkowski, Gruder & Romer, 1983) and, importantly, intergroup behaviors (e.g., Blanz, et al, 1997; Franco & Maass, 1999; Gaertner & Insko, 2001; Paluck, 2009; Pereira, et al., 2009). However, we did not directly investigate the relation between prescriptive norms regarding confrontation of bias and actual responses to bias. Indeed, there is an apparent gap between how positively witnesses' confrontation of bias are evaluated (see Chapter IV; see also Dickter et al., 2012) and the frequency to which witnesses' report confronting bias (Dickter, 2012; Dickter & Newton, 2013). Even though confrontations of biased expressions are socially valued and approved, in Dickter (2012) and Dickter and Newton (2013) studies only about one third of respondents reported confronting those expressions. This discrepancy between prescriptive norms and witnesses' actual responses to bias may be due to several factors. As suggested in Chapter IV, the extent to which confrontations of bias are normative may depend on contextual factors, such as the relationship of the confronter with the person expressing bias – that is, confronting may not always be normative. Accordingly, it is important to consider how other social factors may influence the normativity of confrontation (even though one should bear in mind that norms may only be useful in predicting behavior if they are relatively consistent across social situations). In addition, norms influence behavioral intentions more directly than they influence behaviors (Ajzen, 1987; 1991). Factors that interfere with the translation of intentions to behaviors may also weaken the relationship between prescriptive norms and confrontations of bias. Future research may not only directly investigate the relation of norms and behavior and behavioral intentions in witnesses' confrontation of bias, but also strategies that may facilitate the transition from intentions to confront to actual

confrontations of bias (such as, for example, implementations of intentions; Gollwitzer, 1993; 1999).

In Chapter IV, we showed that threat to the image of a highly-relevant group predicted the prescriptive normativity of witnesses' confrontation of bias. We argued this effect happened because threats to the image of a highly-valued ingroup grants social legitimacy for witnesses to confront bias. However, we did not directly test this hypothesis, a possible future direction for this line of research. In addition, future research may investigate whether threat to the image of a highly-valued ingroup directly predicts witnesses' confrontations of bias. Previous research has shown that a common strategy to deal with threats to the image of the ingroup is to distance oneself from the deviant ingroup member (Brambilla et al., 2013). However, we expect people to be less willing to distance themselves from members of highly-valued ingroups. Alternatively, they may attempt to deal with such image threats by pressuring the deviant ingroup members to conform to the group norms (Marques & Paez, 1994; Sampson & Brandon, 1964). Assuming that the group has egalitarian norms, one way to pressure an ingroup member who express bias to conform to the ingroup norms may be to confront that ingroup member. Future research may test this possibility.

Conclusion

Several factors may influence the extent confrontations of bias are enacted by witnesses, supported by observers and valued by societies. In the current thesis, we showed that who the person expressing bias is in relation to the witness of bias matter for each of these dimensions of witnesses' confrontation of prejudice. Observers have more favorable attitudes when a witness confronts a person who has low power over him or her; witnesses confront an ingroup member less than an outgroup member; and confronting a close person is more normative than confronting a stranger.

Still, the impact of these social factors is not the same for all people. The endorsement and internalization of egalitarian values buffers witnesses against the influence of a shared membership with the person expressing bias; and observers' against the influence of the power asymmetry in the relationship between the person expressing bias and the confronter. That is, people who have strong egalitarian beliefs and standards support and confront even in more unfavorable social conditions, probably because they are strongly motivated to fight for equality and social justice.

Despite considerable progresses in the last decades, we still have a long way to go before we can consider even societies with strong egalitarian norms free of biases. With the current work, we have contributed to the current understanding of factors that may encourage or prevent witnesses' confrontation of bias, a process that may fundamentally challenge public expressions of prejudice.

"A man dies when he refuses to stand up for that which is right. A man dies when he refuses to stand up for justice."

Dr. Martin Luther King Jr., 1965

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage publications.

Ajzen, I. (1987). Attitudes, Traits, and Actions: Dispositional Prediction of Behavior in Personality and Social Psychology. *Advances in Experimental Social Psychology*, *20*(C), 1–63. doi:10.1016/S0065-2601(08)60411-6

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211. doi:10.1016/0749-5978(91)90020-T

Alves, H., & Correia, I. (2008). On the normativity of expressing the belief in a just world: Empirical evidence. *Social Justice Research*, *21*(1), 106–118. doi:10.1007/s11211-007-0060-x

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology*, *60*(2), 241–253. doi:10.1037/0022-3514.60.2.241

Ashburn-Nardo, L., Blanchar, J. C., Petersson, J., Morris, K. A., & Goodwin, S. A. (2014). Do you say something when it's your boss? The role of perpetrator power in prejudice confrontation. *Journal of Social Issues*, *70*(4), 615–636. doi:10.1111/josi.12082

Ayres, M. M., Friedman, C. K., & Leaper, C. (2009). Individual and situational factors related to young women's likelihood of confronting sexism in their everyday lives. *Sex Roles*, *61*(7-8), 449–460. doi:10.1007/s11199-009-9635-3

Becker, J. C., & Barreto, M. (2014). Ways to go: Men's and women's support for aggressive and nonaggressive confrontation of sexism as a function of gender identification. *Journal of Social Issues*, *70*(4), 668–686. doi:10.1111/josi.12085

Becker, J. C., Zawadzki, M. J., & Shields, S. A. (2014). Confronting and reducing sexism: A call for research on intervention. *Journal of Social Issues*, *70*(4), 603–614. doi:10.1111/josi.12081

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*. doi:10.1016/S0065-2601(08)60024-6

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, *79*(6), 993–997. doi:10.1037//0021-9010.79.6.993

Blanchard, F., Lilly, T., & Vaughn, L. (1991). Reducing the expression of racial prejudice. *Psychological Science, 2*(2), 101–105. doi:10.1111/j.1467-9280.1991.tb00108.x

Blanz, M., Mummendey, A., & Otten, S. (1997). Normative evaluations and frequency expectations regarding positive versus negative outcome allocations between groups. *European Journal of Social Psychology, 27*(2), 165–176. doi:10.1002/(SICI)1099-0992(199703)27:2<165::AID-EJSP812>3.0.CO;2-3

Blodorn, A., O'Brien, L. T., & Kordys, J. (2011). Responding to sex-based discrimination: Gender differences in perceived discrimination and implications for legal decision making. *Group Processes & Intergroup Relations*, *15*(3), 409–424. doi:10.1177/1368430211427172

Bornstein, G., Crum, L., Wittenbraker, J., Harring, K., Insko, C. A, & Thibautt, J. (1983). On the measurement of social orientations in the minimal group paradigm. *European Journal of Social Psychology, 13*, 321–350. doi:10.1002/ejsp.2420130402

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*(2), 135–143. doi:10.1002/ejsp.744

Brambilla, M., Sacchi, S., Pagliaro, S., & Ellemers, N. (2013). Morality and intergroup relations: Threats to safety and group image predict the desire to interact with outgroup and ingroup members. *Journal of Experimental Social Psychology*, *49*(5), 811–821. doi:10.1016/j.jesp.2013.04.005

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *51*(1), 149–166. doi:10.1111/j.2044-8309.2010.02011.x

Branscombe, N. R., Wann, D. L., Noel, J. G., & Coleman, J. (1993). In-group or out-group extremity: importance of the threatened social identity. *Personality and Social Psychology Bulletin*, *19*(4), 381–388. doi:10.1177/0146167293194003

Brewer, M. B. (1999). The Psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues, 55*(3), 429–444. doi:10.1111/0022-4537.00126

Brewer, M. B., & Silver, M. (1978). Ingroup bias as a function of task characteristics. *European Journal of Social Psychology*, *8*, 393–400. doi:10.1002/ejsp.2420080312

Brown, R., & Lepore, L. (1996). Prejudice. In A.S.R. Manstead & M. Hewstone (Eds.). *The Blackwell Encyclopedia of Social Psychology*. Blackwell Publishing. Blackwell Reference Online. Retrieved from http://www.blackwellreference.com/subscriber/tocnode.html?id=g9780631202899_chunk_g978063120289918_ss1-17

Burson, A., Crocker, J., & Mischkowski, D. (2012). Two types of value-affirmation: Implications for self-control following social exclusion. *Social Psychological and Personality Science*, *3*(4), 510–516. doi:10.1177/1948550611427773

Cabecinhas, R. (2002). *Racismo e etnicidade em Portugal: Uma análise psicossociológica da homogeneização das minorias*. Doctoral dissertation. Universidade do Minho.

Caprara, G. V., Alessandri, G., & Eisenberg, N. (2012). Prosociality: The contribution of traits, values, and self-efficacy beliefs. *Journal of Personality and Social Psychology*, *102*(6), 1289–1303. doi:10.1037/a0025626

Caprara, G. V., & Steca, P. (2007). Prosocial agency: The contribution of values and self–efficacy beliefs to prosocial behavior across ages. *Journal of Social and Clinical Psychology*, *26*(2), 218–239. doi:10.1521/jscp.2007.26.2.218

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, *24*, 201–234. doi:10.1016/S0065-2601(08)60330-5

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology, 58*(6), 1015–1026. doi:10.1037/0022-3514.58.6.1015

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. *The Handbook of Social Psychology, Vols. 1 and 2*. doi:10.1037/0022-3514.58.6.1015

CICDR - ACIME/ACIDI/ACM (2016). Queixas de discriminação racial CICDR (entre 2000 e 2014). Retrieved from http://www.om.acm.gov.pt/-/discriminacao-de-base-racial-e-etnica.

Clark, M. S., & Taraban, C. (1991). Reactions to and willingness to express emotion in communal and exchange relationships. *Journal of Experimental Social Psychology*, *27*(4), 324–336. doi:10.1016/0022-1031(91)90029-6

CNN/ORC (2015). Poll June 2015. Retrieved from http://i2.cdn.turner.com/cnn/2015/images/06/29/obama.approval.pdf

Cohrs, J. C., Moschner, B., Maes, J., & Kielmann, S. (2005). The motivational bases of right-wing authoritarianism and social dominance orientation : Relations to values and attitudes in the aftermath of September 11, 2001. *Personality & Social Psychology Bulletin*, *31*(10), 1425–1434. doi:10.1177/0146167205275614

Condor, S. (2006). Public prejudice as collaborative accomplishment: Towards a dialogic social psychology of racism. *Journal of Community and Applied Social Psychology*, *16*(1), 1–18. doi:10.1002/casp.845

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359–378. doi:10.1037//0022-3514.82.3.359

Crosby, F. J. (1993). Why complain? *Journal of Social Issues*, *49*(1), 169–184. doi:10.1111/j.1540-4560.1993.tb00916.x

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): reactions to confrontations of racial and gender bias. *Personality & Social Psychology Bulletin*, *29*(4), 532–44. doi:10.1177/0146167202250923

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, *90*(5), 784–803. doi:10.1037/0022-3514.90.5.784

D'Augelli, A. R., & Hershberger, S. L. (1993). African American undergraduates on a predominantly white campus: Academic factors, social networks, and campus climate. *The Journal of Negro Education*, *62*(1), 67–81. doi:10.2307/2295400

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology*, *60*(6), 817–830. doi:10.1037//0022-3514.60.6.8177

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82*(5), 835–848. doi:10.1037/0022-3514.82.5.835

Dickter, C. L. (2012). Confronting hate: Heterosexuals' responses to anti-gay comments. *Journal of Homosexuality*, *59*(8), 1113–30. doi:10.1080/00918369.2012.712817

Dickter, C. L., Kittel, J. A., & Gyurovski, I. I. (2012). Perceptions of non-target confronters in response to racist and heterosexist remarks. *European Journal of Social Psychology, 42*(1), 112–119. doi:10.1002/ejsp.855

Dickter, C. L., & Newton, V. A. (2013). To confront or not to confront: Non-targets' evaluations of and responses to racist comments. *Journal of Applied Social Psychology, 43,* E262–E275. doi:10.1111/jasp.12022

Dodd, E. H., Giuliano, T. A., Boutell, J. M., & Moran, B. E. (2001). Respected or rejected: Perceptions of women who confront sexist remarks. *Sex Roles*, *45*(7-8), 567–577. doi:10.1023/A:1014866915741

Doise, W. (1980). Levels of explanation in the European Journal of Social Psychology. *European Journal of Social Psychology*, 213–231. doi:10.1002/ejsp.2420100302

Doise, W. (1986). *Levels of explanation in social psychology* (E. Mapstone, Trans.). Cambridge: University Press.

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, *11*(4), 315–319. doi:10.1111/1467-9280.00262

Dovidio, J. F. & Gaertner, S. L., (2010). Intergroup bias. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (eds.), *Handbook of Social Psychology* (5th Edition, pp. 1084-1121). Hoboken, N.J: John Wiley & Sons.

Drury, B. J., & Kaiser, C. R. (2014). Allies against sexism: The role of men in confronting sexism. *Journal of Social Issues, 70*(4), 637–652. doi:10.1111/josi.12083

Dubois, N. (2003). Introduction: The concept of norm. In N. Dubois (Ed.), *A sociocognitive approach to social norms*, (pp 1-16). London: Routledge.

Duriez, B., & Van Hiel, A. (2002). The march of modern fascism. A comparison of social dominance orientation and authoritarianism. *Personality and Individual Differences*, *32*(7), 1199–1213. doi:10.1016/S0191-8869(01)00086-1

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology*, *36*(4), 424–435. doi:10.1037//0022-3514.36.4.424

Eccleston, C. P., & Major, B. N. (2006). Attributions to discrimination and self-esteem: The role of group identification and appraisals. *Group Processes & Intergroup Relations*, *9*(2), 147–162. doi:10.1177/1368430206062074

Effron, D. A., Miller, D. T., & Monin, B. (2012). Inventing racist roads not taken: The licensing effect of immoral counterfactual behaviors. *Journal of Personality and Social Psychology*, *103*(6), 916–932. doi:10.1037/a0030008

Effron, D. A., & Miller, D. T. (2012). How the moralization of issues grants social legitimacy to act on one's attitudes. *Personality & Social Psychology Bulletin*, *38*(5), 690–701. doi:10.1177/0146167211435982

Eliezer, D., & Major, B. (2012). It's not your fault: The social costs of claiming discrimination on behalf of someone else. *Group Processes & Intergroup Relations*, *15*(4), 487–502. doi:10.1177/1368430211432894

Essed, P. (1991) *Understanding Everyday Racism: An Interdisciplinary Theory*. Newbury Park, CA: Sage.

European Union Agency for Fundamental Rights. (2009). *EU-MIDIS. European Union Minorities and Discrimination Survey. Main Results Report*. doi:10.2811/32815

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi:10.3758/BF03193146

Fazio, R. H. (1990). Multiple Processes by which Attitudes Guide Behavior: The Mode Model as an Integrative Framework. *Advances in Experimental Social Psychology*, *23*(C), 75–109. doi:10.1016/S0065-2601(08)60318-4

Feagin, J. R. (1991). The continuing significance of race: Antiblack discrimination in public places. *American Sociological Review*, *56*(1), 101. doi:10.2307/2095676

Feather, N. T. (1995). Values, valences, and choice: The influences of values on the perceived attractiveness and choice of alternatives. *Journal of Personality and Social Psychology*, *68*(6), 1135–1151. doi:10.1037/0022-3514.68.6.1135

Feather, N. T., & McKee, I. R. (2008). Values and prejudice: Predictors of attitudes towards Australian Aborigines. *Australian Journal of Psychology*, *60*(2), 80–90. doi:10.1080/00049530701449513

Feather, N. T., & McKee, I. R. (2012). Values, right-wing authoritarianism, social dominance orientation, and ambivalent attitudes toward women. *Journal of Applied Social Psychology*, *42*, 2479–2504. doi:10.1111/j.1559-1816.2012.00950.x

Fonseca, M. L., & McGarrigle, J. (2012) *Policy recommendations: Promoting interethnic coexistence, social cohesion and reducing anti-immigrant attitudes (GEITONIES project).* Lisboa, Centro de Estudos Geográficos (MIGRARE Working Paper Series # 7), p. 19.

Franco, F. M., & Maass, A. (1999). Intentional control over prejudice: When the choice of the measure matters. *European Journal of Social Psychology*, *29*(4), 469–477. doi:10.1002/(SICI)1099-0992(199906)29:4<469::AID-EJSP938>3.0.CO;2-S

Gaertner, L., & Insko, C. A. (2001). On the measurement of social orientations in the minimal group paradigm: Norms as moderators of the expression of intergroup bias. *European Journal of Social Psychology*, *31*(2), 143–154. doi:10.1002/ejsp.28

Garcia, D. M., Reser, A. H., Amo, R. B., Redersdorff, S., & Branscombe, N. R. (2005). Perceivers' responses to in-group and out-group members who blame a negative outcome on discrimination. *Personality & Social Psychology Bulletin*, *31*(6), 769–80. doi:10.1177/0146167204271584

Gilibert, D., & Cambon, L. (2003). Paradigms of the sociocognitive approach. In N. Dubois (Ed.), *A Sociocognitve Approach to Social Norms* (pp. 38-69). London: Routledge.

Gollwitzer, P. M. (1993). Goal achievement: The role of intentions. *European Review of Social Psychology*, *4*(1), 141–185. doi:10.1080/14792779343000059

Gollwitzer, P. M. (1999). Implementation intentions. *American Psychologist*, *54*(7), 493–503. doi:10.1177/0146167207311201

Good, J., Moss-Racusin, C., & Sanchez, D. (2012). When do we confront? Perceptions of costs and benefits predict confronting discrimination on behalf of the self and others. *Psychology of Women Quarterly*, *36*(2), 210–226. doi:10.1177/0361684312440958

Gruder, C. L., Romer, D., & Korth, D. (1978). Dependency and fault as determinants of helping. *Journal of Experimental Social Psychology*, *14*(2), 227–235. doi:10.1016/0022-1031(78)90028-8

Gulker, J. E., Mark, A. Y., & Monteith, M. J. (2013). Confronting prejudice: The who, what, and why of confrontation effectiveness. *Social Influence*, *8*(4), 280–293. doi:10.1080/15534510.2012.736879

Hayes, A. F. (2013). *Introduction to mediation, moderation and conditional process analysis.* New York, New York: Guilford Press.

Hewstone, M. (1990). The 'ultimate attribution error'? A review of the literature on intergroup causal attribution. European Journal of Social Psychology, 20(4), 311–335. doi:10.1002/ejsp.2420200404

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*(1), 575–604. doi:10.1146/annurev.psych.53.100901.135109

Hewstone, M., & Ward, C. (1985). Ethnocentrism and causal attribution in Southeast Asia. *Journal of Personality and Social Psychology*, *48*(3), 614–623. doi:10.1037/0022-3514.48.3.614

Holahan, C. J. (1977). Effects of urban size and heterogeneity on judged appropriateness of altruistic responses: Situational vs. subject variables. *Sociometry*, *40*(4), 378–383. doi:10.2307/3033488

Homer, P. M., & Kahle, L. R. (1988). A structureal equation test of the value-attitude-behavior hierarchy. *Journal of Personality and Social Psychology*, *54*(4), 638–646. doi:10.1037/0022-3514.54.4.638

Huddleston, T., Bilgili, O., Joki, A., & Vankova, Z. (2015). *Migrant integration policy index 2015*. Barcelona/Brussels: CIDOB and MPG | www.mipex.eu.

Hyers, L. L. (2007). Resisting prejudice every day: Exploring women's assertive responses to anti-black racism, anti-semitism, heterosexism, and sexism. *Sex Roles*, *56*(1-2), 1–12. doi:10.1007/s11199-006-9142-8

Inman, M. L., & Baron, R. S. (1996). Influence of prototypes on perceptions of prejudice. *Journal of Personality and Social Psychology*, *70*(4), 727–39. doi:10.1037/0022-3514.70.4.727

Jellison, J. M., & Green, J. (1981). A self-presentation approach to the fundamental attribution error: The norm of internality. *Journal of Personality and Social Psychology*, *40*(4), 643–649. doi: 10.1037//0022-3514.40.4.643

Johns, M., Schmader, T., & Lickel, B. (2005). Ashamed to be an American? The role of identification in predicting vicarious shame for anti-Arab prejudice after 9–11. *Self and Identity*, *4*(4), 331–348. doi:10.1080/15298860500145822

Kahn, K. B., Barreto, M., Kaiser, C. R., & Rego, M. S. (2015). When do high and low status group members support confrontation? The role of perceived pervasiveness of prejudice. *British Journal of Social Psychology*, Online first. doi:10.1111/bjso.12117

Kaiser, C. R., Hagiwara, N., Malahy, L. W., & Wilkins, C. L. (2009). Group identification moderates attitudes toward ingroup members who confront discrimination. *Journal of Experimental Social Psychology*, *45*(4), 770–777. doi:10.1016/j.jesp.2009.04.027

Kaiser, C. R., & Miller, C. T. (2001). Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin*, *27*(2), 254–263. doi:10.1177/0146167201272010

Kaiser, C. R., & Miller, C. T. (2003). Derogating the victim: The interpersonal consequences of blaming events on discrimination. *Group Processes & Intergroup Relations*, *6*(3), 227–237. doi:10.1037/e633872013-490

Kaiser, C. R., & Miller, C. T. (2004). A stress and coping perspective on confronting abstract sexism. *Psychology of Women Quarterly*, *28*, 168–178. doi:10.1111/j.1471-6402.2004.00133.x

Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, *26*(8), 1002–1012. doi:10.1177/01461672002610009

Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology*, *55*(6), 893–905. doi:10.1037/0022-3514.55.6.893

Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and behavioral responses to racism. *Science*, *323*, 276–278. doi:10.1126/science.1164951

Kay, A. C., & Jost, J. T. (2003). Complementary justice: Effects of "poor but happy" and "poor but honest" stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology, 85*, 823–837. doi:10.1037/0022-3514.85.5.823

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review*, *110*(2), 265–284. doi:10.1037/0033-295X.110.2.265

Kuntz, A., Davidov, E., Schwartz, S. H., & Schmidt, P. (2015). Human values, legal regulation, and approval of homosexuality in Europe : A cross- country comparison. *European Journal of Social Psychology*, *45*, 120–134. doi:10.1002/ejsp.2068

Lavado, S., Pereira, C. R., Dovidio, J. F., & Vala, J. (2016). Evaluations of witnesses' responses to bias: Universalism-concern and the costs of confrontation. *Personality and Individual Differences*, *96*, 172–180. doi:10.1016/j.paid.2016.03.003

Lee, E. A., Soto, J. A., Swim, J. K., & Bernstein, M. J. (2012). Bitter reproach or sweet revenge: Cultural differences in response to racism. *Personality and Social Psychology Bulletin*, *38*(7), 920–938. doi:10.1177/0146167212440292

Lickel, B., Hamilton, D. L., Wieczorkowska, G., Lewis, A., Sherman, S. J., & Uhles, A. N. (2000). Varieties of groups and the perception of group entitativity. *Journal of Personality and Social Psychology*, *78*(2), 223–246. doi:10.1037/0022-3514.78.2.223

Lickel, B., Schmader, T., Curtis, M., Scarnier, M., & Ames, D. R. (2005). Vicarious shame and guilt. *Group Processes & Intergroup Relations*, *8*(2), 145–157. doi:10.1177/1368430205051064

Lima-Nunes, A., Pereira, C. R., & Correia, I. (2013). Restricting the scope of justice to justify discrimination: The role played by justice perceptions in discrimination against immigrants. *European Journal of Social Psychology*, *43*(7), 627–636. doi:10.1002/ejsp.1981

Lönnqvist, J.-E., Leikas, S., Paunonen, S., Nissinen, V., & Verkasalo, M. (2006). Conformism moderates de relations between values, anticipated regret and behavior. *Personality and Social Psychology Bulletin*, *32*(11), 1469–1481. doi:10.1177/0146167206291672

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology*, *57*(6), 981–993. doi:10.1037/0022-3514.57.6.981

Mallett, R. K., & Melchiori, K. J. (2014). Goal preference shapes confrontations of sexism. *Personality & Social Psychology Bulletin*, *40*(5), 646–56. doi:10.1177/0146167214521468

Mallett, R. K., & Wagner, D. E. (2011). The unexpectedly positive consequences of confronting sexism. *Journal of Experimental Social Psychology*, *47*(1), 215–220. doi:10.1016/j.jesp.2010.10.001

Marques, J. M., & Paez, D. (1994). The "black sheep effect": Social categorization, rejection of ingroup deviates, and perception of group variability. *European Review of Social Psychology*. doi:10.1080/14792779543000011

Mendes, M. M., & Candeias, P. (2013). Immigrant perceptions of ethnic and racial discrimination: patterns and singularities in a municipality in the Lisbon Metropolitan Area. *International Review of Sociology*, *23*(2), 1–20. doi:10.1080/03906701.2013.804299

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, *80*(4), 252–283. doi:10.1037/h0035002

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*(3), 469–485. doi:10.1037/0022-3514.65.3.469

Moreland, R. L., & Levine, J. M. (1982). Socialization in small groups: Temporal changes in individual-group relations. *Advances in Experimental Social Psychology*, *15*, 137–192. doi:10.1016/S0065-2601(08)60297-X

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup bias as a function of salience, relevance, and status - an integration. *European Journal of Social Psychology*, *22*, 103–122. doi:10.1002/ejsp.2420220202

Mummendey, A., & Wenzel, M. (1999). Social discrimination and tolerance in intergroup relations: Reactions to intergroup difference. *Personality and Social Psychology Review*, *3*(2), 158–174. doi: 10.1207/s15327957pspr0302_4

Neto, F. (2001). Satisfaction with life among adolescents from immigrant families in Portugal. *Journal of Youth and Adolescence*, *30*(1), 53–67. doi:10.1023/A:1005272805052

Neto, F. (2006). Psycho-social predictors of perceived discrimination among adolescents of immigrant background: A Portuguese study. *Journal of Ethnic and Migration Studies*, *32*(April 2013), 89–109. doi:10.1080/13691830500335507

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259. doi:10.1037/0033-295X.84.3.231

Oishi, S., Kesebir, S., & Diener, E. (2011). Income inequality and happiness. *Psychological Science*, *22*(9), 1095–1100. doi:10.1177/0956797611417262

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict with the mass media: A field experiment in Rwanda. *Journal of Personality and Social Psychology*, *96*(3), 574–587. doi:10.1037/a0011989

Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., … Gee, G. (2015). Racism as a Determinant of Health: A Systematic Review and Meta-Analysis. *Plos One*, *10*(9), e0138511. doi:10.1371/journal.pone.0138511

Pascoe, E. A., & Richman, L. S. (2009). Perceived discrimination and health: A meta-analytic review. *Psychological Bulletin*, *135*(4), 531–554. doi:10.1037/a0016059

Peixe, B., Rosario, E., Silva, E., Soares, P., Kumar, R., Ralha, T., & Santos, T. (2008). O racismo e xenofobia em Portugal (2001-2007). Retrieved from http://www.amnistia-internacional.pt/files/Estudo_Racismo_Portugal.pdf

Penner, L. A., Dovidio, J. F., Piliavin, J. A., & Schroeder, D. A. (2005). Prosocial behavior: Multi-level perspectives. *Annual Review of Psychology*, *56*, 365–392. doi:10.1146/annurev.psych.56.091103.070141

Peral, E. B., & Ramos, A. (2014). Neighbours: Determinants of whom Europeans want to keep a distance. In Arts, W., Halman, L. (Eds.), *Value contrasts and consensus in present-day Europe: painting Europe's moral landscapes* (pp. 117-141). Leiden: Brill

Pereira, C., Vala, J., & Costa-Lopes, R. (2010). From prejudice to discrimination: The legitimizing role of perceived threat in discrimination against immigrants. *European Journal of Social Psychology*, *40*, 1231–1250. doi:10.1002/ejsp

Pereira, C., Vala, J., & Leyens, J. P. (2009). From infra-humanization to discrimination: The mediation of symbolic threat needs egalitarian norms. *Journal of Experimental Social Psychology*, *45*(2), 336–344. doi:10.1016/j.jesp.2008.10.010

Perry, S. P., Murphy, M. C., & Dovidio, J. F. (2015). Modern prejudice: Subtle, but unconscious? The role of bias awareness in Whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology*, *61*, 64–78. doi:10.1016/j.jesp.2015.06.007

Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and Social Psychology Bulletin*, *5*(4), 461–476. doi:0803973233

Petty, R. E., Fleming, M. A., Priester, J. R., & Feinstein, A. H. (2001). Individual versus group interest violation: Surprise as a determinant of argument scrutiny and persuasion. *Social Cognition*, *19*(4), 418–442. doi:10.1521/soco.19.4.418.20758

Pew Research Center (2015). *The evolving role of news on twitter and facebook*. Retrieved from http://www.journalism.org/2015/07/14/the-evolving-role-of-news-on-twitter-and-facebook/

Piliavin, J. A., Dovidio, J. F., Gaertner, S. L., & Clark, R. D., III. (1981). *Emergency intervention*. New York: Academic Press.

Pinto, I. R., Marques, J. M., & Paez, D. (2015). National identification as a function of perceived social control: A subjective group dynamics analysis. *Group Processes & Intergroup Relations*, (146). doi:10.1177/1368430215577225

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*(3), 811–832. doi:10.1037/0022-3514.75.3.811

Plant, E. A., & Devine, P. G. (2009). The active control of prejudice: Unpacking the intentions guiding control efforts. *Journal of Personality and Social Psychology*, *96*(3), 640–652. doi:10.1037/a0012960

Plant, E. A., Devine, P. G., & Brazy, P. C. (2003). The bogus pipeline and motivations to respond without prejudice: Revisiting the fading and faking of racial prejudice. *Group Processes & Intergroup Relations*, *6*(2), 187–200. doi:10.1177/1368430203006002004

Plant, E. A., Devine, P. G., & Peruche, M. B. (2010). Routes to positive interracial interactions: approaching egalitarianism or avoiding prejudice. *Personality and Social Psychology Bulletin*, *36*(9), 1135–1147. doi:10.1177/0146167210378018

Pronin, E. (2009). The introspection illusion. *Advances in Experimental Social Psychology*, *41*, 1–67. doi:10.1016/S0065-2601(08)00401-2

Ramos, A. (2011). Human values and opposition towards immigration in Europe. Doctoral thesis. University of Lisbon, Institute of Social Sciences.

Ramos, A., Pereira, C. R., & Vala, J. (2016). Economic crisis, human values and attitudes towards immigrants. In: M. Voicu, I. C. Mochmann & H. Dülmer (editors), *Values, economic crisis and democracy* (pp. 104-137). New York: Routledge.

Rasinski, H. M., & Czopp, A. M. (2010). The effect of target status on witnesses' reactions to confrontations of bias. *Basic and Applied Social Psychology*, *32*(1), 8–16. doi:10.1080/01973530903539754

Rattan, A., & Dweck, C. S. (2010). Who confronts prejudice?: The role of implicit theories in the motivation to confront prejudice. *Psychological Science*, *21*(7), 952–959. doi:10.1177/0956797610374740

Rokeach, M. (1973). *The nature of human values*. New York, NY: Free Press.

Rutkowski, G. K., Gruder, C. L., & Romer, D. (1983). Group cohesiveness, social norms, and bystander intervention. *Journal of Personality and Social Psychology*, *44*(3), 545–552. doi:10.1037/0022-3514.44.3.545

Rutland, A., Killen, M., & Abrams, D. (2010). A new social-cognitive developmental perspective on prejudice: The interplay between morality and group identity. *Perspectives on*, *5*(3), 279–291. doi:10.1177/1745691610369468

Salvatore, J., & Shelton, J. N. (2007). Cognitive costs of exposure to racial prejudice. *Psychological Science*, *18*(9), 810–5. doi:10.1111/j.1467-9280.2007.01984.x

Sampson, E. E., & Brandon, A. C. (1964). The effects of role and opinion deviation on small group behavior. *Sociometry*, *27*(3), 261–281. doi:10.2307/2785618

Santos, T., Reis Oliveira, C., Rosário, E., Kumar, R., & Brigadeiro, E. (2005). *Research survey on migrants' experiences of racism and discrimination in Portugal*. Oeiras: Númena.

Schmader, T., Croft, A., Scarnier, M., Lickel, B., & Mendes, W. B. (2012). Implicit and explicit emotional reactions to witnessing prejudice. *Group Processes & Intergroup Relations*, *15*(3), 379–392. doi:10.1177/1368430211426163

Schmitt, M. T., Branscombe, N. R., Postmes, T., & Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: A meta-analytic review. *Psychological Bulletin*, *140*(4), 921–948. doi:10.1037/a0035754

Schwartz, S. H. (1973). Normative explanations of helping behavior: A critique, proposal, and empirical test. *Journal of Experimental Social Psychology*, *9*(4), 349–364. doi:10.1016/0022-1031(73)90071-1

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1–65. doi:10.1016/S0065-2601(08)60281-6

Schwartz, S. H. (1994). Are there universal aspects in the structure and content of human values? *Journal of Social Issues, 50*(4), 19–45.

Schwartz, S. H. (2010). Basic values: How they motivate and inhibit prosocial behavior. In M. Mikulincer & P. R. Shaver (Eds.) *Prosocial motives, emotions, and behavior: The better angels of our nature* (pp. 221–241). Washington DC: American Psychological Association. doi:10.1037/12061-012

Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture, 2,* 1–20. doi:10.9707/2307-0919.1116

Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology, 53*(3), 550–562. doi:10.1037//0022-3514.53.3.550

Schwartz, S. H., & Butenko, T. (2014). Values and behavior: Validating the refined values theory in Russia. *European Journal of Social Psychology, 44,* 799–813. doi:10.1002/ejsp.2053

Schwartz, S. H., Caprara, G. V., & Vecchione, M. (2010). Basic personal values, core political values, and voting: A longitudinal analysis. *Political Psychology, 31*(3), 421–452. doi:10.1111/j.1467-9221.2010.00764.x

Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., … Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology, 103*(4), 663–688. doi:10.1037/a0029393

Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., Harris, M., & Owens, V. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology, 32*(5), 519–542. doi:10.1177/0022022101032005001

Shelton, J. N., & Stewart, R. E. (2004). Confronting perpetrators of prejudice: The inhibitory effects of social costs. *Psychology of Women Quarterly, 28*(3), 215–223. doi:10.1111/j.1471-6402.2004.00138.x

Shelton, J. N., Richeson, J. A., Salvatore, J., & Hill, D. M. (2006). Silence is not golden: Intrapersonal consequences of not challenging prejudice. In S. Levin & C. Van Laar (Eds.), *Stigma and group inequality: Social psychological perspectives* (pp. 65–81). Mahwah, NJ: Erlbaum.

Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey and E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 883-947). New York, NY: Random House.

Sommers, S. R., & Norton, M. I. (2006). Lay theories about White racists: What constitutes racism (and what doesn't). *Group Processes & Intergroup Relations, 9*(1), 117–138. doi:10.1177/1368430206059881

Stangor, C. G. (1996). Stereotyping. In A.S.R. Manstead & M. Hewstone (eds.). *The Blackwell Encyclopedia of Social Psychology*. Blackwell Publishing. Blackwell Reference Online. Retrieved from http://www.blackwellreference.com/subscriber/tocnode.html?id=g9780631202899_chunk_g978063120289921_ss1-53

Swim, J. K., Gervais, S. J., Pearson, N., & Stangor, C. (2009). Managing the message: Using social influence and attitude change strategies to confront interpersonal discrimination. In F. Butera & J. M. Levine (Eds.), *Coping with minority status. Responses to exclusion and inclusion.* New York, NY: Cambridge University Press.

Swim, J. K., & Hyers, L. L. (1999). Excuse Me—What did you just say?!: Women's public and private responses to sexist remarks. *Journal of Experimental Social Psychology*, *35*(1), 68–88. doi:10.1006/jesp.1998.1370

Swim, J. K., Hyers, L. L., Cohen, L. L., & Ferguson, M. J. (2001). Everyday Sexism: Evidence for Its Incidence, Nature, and Psychological Impact From Three Daily Diary Studies. *Journal of Social Issues*, *57*(1), 31–53. doi:10.1111/0022-4537.00200

Swim, J. K., Hyers, L. L., Cohen, L. L., Fitzgerald, D. C., & Bylsma, W. H. (2003). African American college students' experiences with everyday racism: Characteristics of and responses to these incidents. *Journal of Black Psychology*, *29*(1), 38–67. doi:10.1177/0095798402239228

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, *223*(5), 96–102. doi:10.1038/scientificamerican1170-96

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*, 1–39. doi:http://dx.doi.org/10.1146/annurev.ps.33.020182.000245

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*(2), 149–178. doi:10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S.Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Monterey, CA: Brooks/Cole.

Tyler, T. R. & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review*, *7*(4), 349–361. doi:10.1207/S15327957PSPR0704

Vala, J., Brito, R. & Lopes, D. (1999/2015). *Expressões dos racismos em Portugal. 2ª edição online*. Lisboa: ICS. Imprensa de Ciências Sociais.

Vala, J., Lopes, D. & Lima, M. (2008). Black Immigrants in Portugal: Luso – Tropicalism and Prejudice. *Journal of Social Issues*, *64*(2), 287–302. doi:10.1111/j.1540-4560.2008.00562.x

Vala, J. & Pereira, C. R. (2012). Racism: An evolving virus. In F. Bethencourt & A. J. Pearce (Eds.), *Racism and ethnic relations in the portuguese-speaking world* (pp. 49-70). New York: Oxford University Press.

Van Dijk, T. A. (1992). Discourse and the denial of racism. *Discourse & Society*, *3*(1), 87–118. doi:10.1177/0957926592003001005

Verplanken, B. & Holland, R. W. (2002). Motivated decision making: Effects of activation and self-centrality of values on choices and behavior. *Journal of Personality and Social Psychology*, *82*(3), 434–447. doi:10.1037/0022-3514.82.3.434

Waldzus, S., & Mummendey, A. (2004). Inclusion in a superordinate category, in-group prototypicality, and attitudes towards out-groups. *Journal of Experimental Social Psychology*, *40*(4), 466–477. doi:10.1016/j.jesp.2003.09.003

Wang, K., Silverman, A., Gwinn, J. D., & Dovidio, J. F. (2014). Independent or ungrateful? Consequences of confronting patronizing help for people with disabilities. *Group Processes & Intergroup Relations*. doi:10.1177/1368430214550345

Weber, J. G. (1994). The nature of ethnocentric attribution bias: Ingroup protection or enhancement? Journal of Experimental Social Psychology. doi:10.1006/jesp.1994.1023

Winslow, M. P. (2004). Reactions to the imputation of prejudice. *Basic and Applied Social Psychology*, *26*(4), 289–297. doi:10.1207/s15324834basp2604_5

Woodzicka, J. A. & LaFrance, M. (2001). Real versus imagined gender harassment. *Journal of Social Issues, 57*(1), 15–30. doi:10.1111/0022-4537.00199

Appendixes

Appendix A

Materials used in Study 1 and Study 2 of Chapter II

1. Scenario used in Study 1 and Study 2. The low cost/low power condition is in brackets.

Vamos pedir-lhe que leia uma situação que ocorreu recentemente numa entrevista de emprego em Portugal. Por favor, tente imaginar esta situação o mais realisticamente que conseguir.

Leia por favor a descrição da situação com atenção, uma vez que as perguntas que lhe vamos fazer se referem aos eventos e personagem da história.

O Paulo está numa sala à espera de ser chamado para uma entrevista de emprego. Esta entrevista [não] é muito importante para o Paulo, uma vez que esta é a sua terceira entrevista de emprego deste mês e [ainda não teve nenhuma oferta de trabalho] já recebeu algumas ofertas de trabalho interessantes. O Paulo [não] tem de dar o seu melhor nesta entrevista, porque [não] precisa mesmo deste emprego. Além disso, o cargo [não] parece muito interessante e ele [não quer assim tanto] quer mesmo a oportunidade de trabalhar nesta empresa.

O Paulo é chamado para a entrevista. É cumprimentado pelo entrevistador, um homem alto e loiro que usa fato. Ambos se sentam e o entrevistador começa a fazer perguntas ao Paulo. O Paulo tem a impressão que a entrevista está a correr bem. No final, o entrevistador dá um aperto de mão ao Paulo e diz "Gostei muito de si e acho que o seu perfil se adequa à nossa empresa. Tivemos bastantes candidatos negros, por isso é bom ter alguém branco, para variar. Entrarei em contacto consigo quando tivermos uma decisão."

Confrontation condition: O Paulo decide responder ao entrevistador. Ele diz "Acho que a cor da pele não devia ter nada a ver com isto."

Non-confrontation condition: O Paulo decide responder ao entrevistador. Ele diz "Fico então a aguardar o seu contacto".

2. Attitudes toward Paulo's behavior (Study 1 and Study 2)

Em que medida acha que a reação do Paulo foi…

	Nada			Um pouco			Muito
Apropriada	1	2	3	4	5	6	7
Irracional	1	2	3	4	5	6	7
Sensata	1	2	3	4	5	6	7
Despropositada	1	2	3	4	5	6	7

3. Manipulation checks (Study 2)

Quão importante era para o Paulo conseguir o emprego?

Nada			Um pouco			Muito
1	2	3	4	5	6	7

Qual a probabilidade de o Paulo conseguir o emprego?

Muito improvável			Nem improvável nem provável			Muito provável
1	2	3	4	5	6	7

4. PVQ-5X (Schwartz et al., 2012) – Female version (Study 2)

Vamos apresentar-lhe descrições breves de algumas pessoas. Por favor leia cada uma das descrições e pense em que medida cada uma dessas pessoas é ou não parecida consigo. Assinale com um círculo o número que melhor indicar em que medida a pessoa descrita é parecida consigo. Tente, por favor, responder a todas as questões.

… Quão parecida consigo é esta pessoa?

	Exacta-mente como eu	Muito parecido comigo	Parecido comigo	Um bocadinho parecido comigo	Nada parecido comigo	Não tem nada a ver comigo
Uma mulher para quem é importante proteger os elementos mais frágeis e vulneráveis da sociedade.	1	2	3	4	5	6
Uma mulher que quer que as pessoas façam o que ela diz.	1	2	3	4	5	6
Uma mulher que acredita seriamente que deve cuidar da natureza.	1	2	3	4	5	6
Uma mulher para quem é importante ser leal para com as pessoas que lhe são próximas.	1	2	3	4	5	6
Uma mulher para quem é importante a sensação de poder que o dinheiro pode dar.	1	2	3	4	5	6
Uma mulher que pensa que é importante ser ambiciosa.	1	2	3	4	5	6
Uma mulher para quem é importante agir contra as ameaças ao mundo natural.	1	2	3	4	5	6
Uma mulher para quem é importante ser rica.	1	2	3	4	5	6
Uma mulher para quem é muito importante ajudar as pessoas que lhe são queridas.	1	2	3	4	5	6
Uma mulher que acha que todas as pessoas no mundo devem ter as mesmas oportunidades na vida.	1	2	3	4	5	6
Uma mulher para quem é importante cuidar do bem-estar das pessoas que lhe são próximas.	1	2	3	4	5	6
Uma mulher para quem é importante ser ela a dizer aos outros o que devem fazer.	1	2	3	4	5	6
Uma mulher para quem é importante ouvir pessoas diferentes de si.	1	2	3	4	5	6
Uma mulher para quem é importante ter muito sucesso.	1	2	3	4	5	6
Uma mulher que se esforça por ser uma amiga de confiança e com quem se pode contar sempre.	1	2	3	4	5	6
Uma mulher para quem é importante proteger o ambiente da destruição ou poluição.	1	2	3	4	5	6
Uma mulher que quer que todas as pessoas sejam tratadas de forma justa, mesmo aquelas que ela não conhece.	1	2	3	4	5	6
Uma mulher que quer ser admirada pelo que consegue realizar.	1	2	3	4	5	6
Uma mulher que quer que as pessoas com quem costuma estar possam contar com ela plenamente.	1	2	3	4	5	6

Uma mulher para quem é importante compreender as outras pessoas, mesmo quando não concorda com elas.	1	2	3	4	5	6

5. PVQ-5X (Schwartz et al., 2012) – Male version (Study 2)

Vamos apresentar-lhe descrições breves de algumas pessoas. Por favor leia cada uma das descrições e pense em que medida cada uma dessas pessoas é ou não parecida consigo. Assinale com um círculo o número que melhor indicar em que medida a pessoa descrita é parecida consigo. Tente, por favor, responder a todas as questões.

… Quão parecida consigo é esta pessoa?

	Exacta-mente como eu	Muito parecido comigo	Parecido comigo	Um bocadinho parecido comigo	Nada parecido comigo	Não tem nada a ver comigo
Um homem para quem é importante proteger os elementos mais frágeis e vulneráveis da sociedade.	1	2	3	4	5	6
Um homem que quer que as pessoas façam o que ele diz.	1	2	3	4	5	6
Um homem que acredita seriamente que deve cuidar da natureza.	1	2	3	4	5	6
Um homem para quem é importante ser leal para com as pessoas que lhe são próximas.	1	2	3	4	5	6
Um homem para quem é importante a sensação de poder que o dinheiro pode dar.	1	2	3	4	5	6
Um homem que pensa que é importante ser ambicioso.	1	2	3	4	5	6
Um homem para quem é importante agir contra as ameaças ao mundo natural.	1	2	3	4	5	6
Um homem para quem é importante ser rico.	1	2	3	4	5	6
Um homem para quem é muito importante ajudar as pessoas que lhe são queridas.	1	2	3	4	5	6
Um homem que acha que todas as pessoas no mundo devem ter as mesmas oportunidades na vida.	1	2	3	4	5	6
Um homem para quem é importante cuidar do bem-estar das pessoas que lhe são próximas.	1	2	3	4	5	6
Um homem para quem é importante ser ele a dizer aos outros o que devem fazer.	1	2	3	4	5	6
Um homem para quem é importante ouvir pessoas diferentes de si.	1	2	3	4	5	6
Um homem para quem é importante ter muito sucesso.	1	2	3	4	5	6
Um homem que se esforça por ser um amigo de confiança e com quem se pode contar sempre.	1	2	3	4	5	6
Um homem para quem é importante proteger o ambiente da destruição ou poluição.	1	2	3	4	5	6
Um homem que quer que todas as pessoas sejam tratadas de forma justa, mesmo aquelas que ele não conhece.	1	2	3	4	5	6

Um homem que quer ser admirado pelo que consegue realizar.	1	2	3	4	5	6
Um homem que quer que as pessoas com quem costuma estar possam contar com ele plenamente.	1	2	3	4	5	6
Um homem para quem é importante compreender as outras pessoas, mesmo quando não concorda com elas.	1	2	3	4	5	6

6. Manipulation check (Study 2)

Qual foi a reação do Paulo ao comentário do entrevistador?

a) Não disse nada.
b) Disse "Acho que a cor da pele não devia ter nada a ver com isto."
c) Concordou com o comentário do entrevistador.
d) Disse "Fico então a aguardar o seu contacto".

Appendix B

Materials used in Study 1 and Study 2 of Chapter III

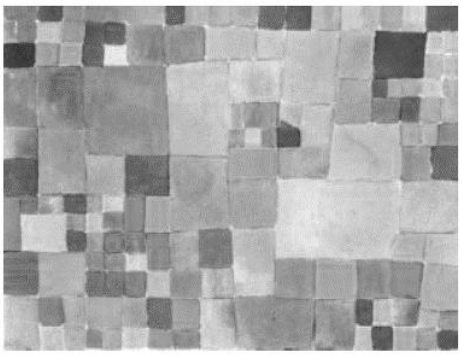
1. Minimal group paradigm (Study 1 and Study 2)

As part of your profile that will be shared with the other MTurker, we will conduct a Artistic Preference Test.

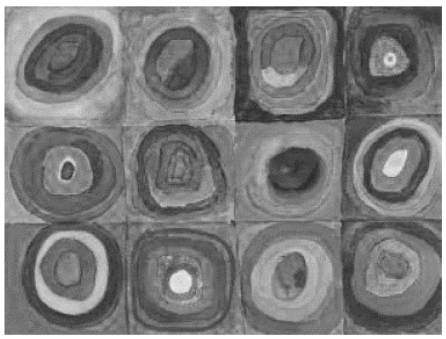
You will see pairs of pictures painted by two modern foreign painters. In each pair, there will be a picture painted by Klee and a picture painted by Kandinsky. You will be asked to indicate which picture you prefer.

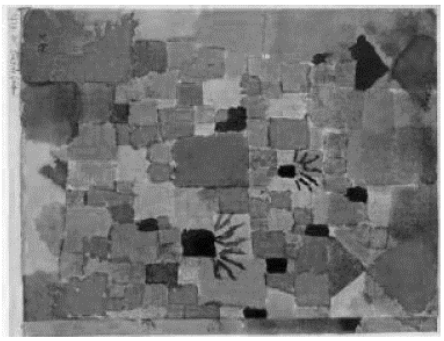
Click next when you are ready to start.

Which one of these pictures do you prefer?

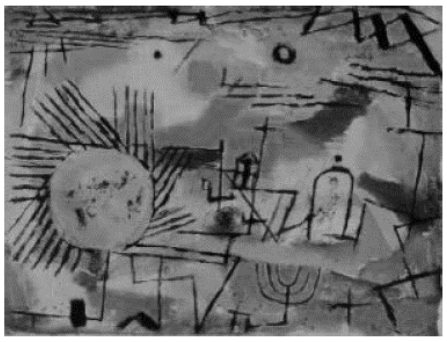


Which one of these pictures do you prefer?

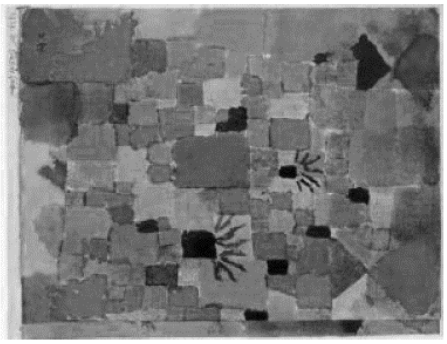




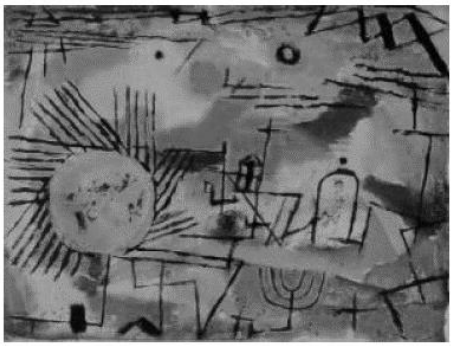
Which one of these pictures do you prefer?

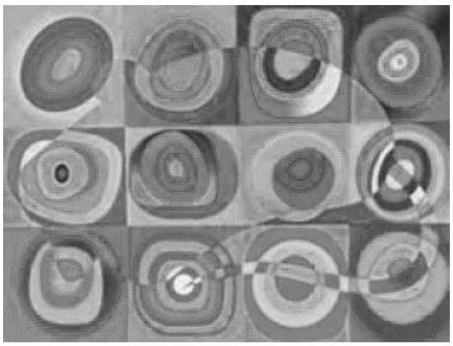


Which one of these pictures do you prefer?

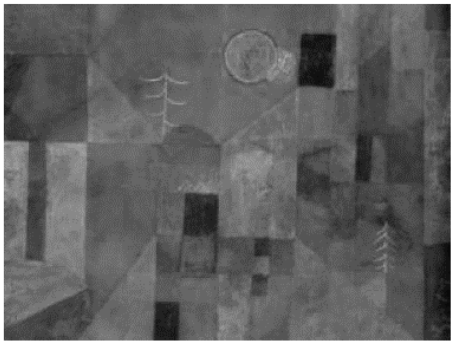


Which one of these pictures do you prefer?

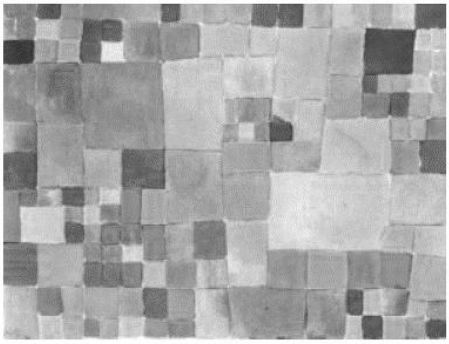


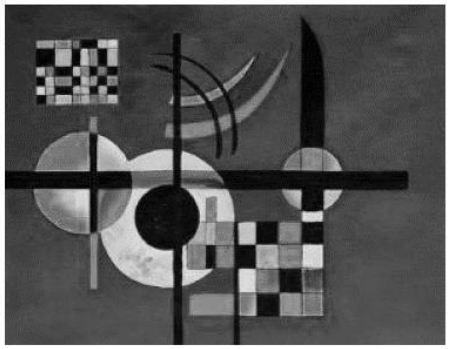


Which one of these pictures do you prefer?



Which one of these pictures do you prefer?





2. Minimal group paradigm: Results (Klee[22]) (Study 1 and Study 2)

In this test, some people consistently prefer the paintings of Klee (i.e., they belong to the Klees) and some people consistently prefer the paintings of Kandinsky (i.e., they belong to the Kandinskys).

Being a member of the Klees or of the Kandinskys has been shown to relate to personality variables, such as general outlook in life and sociability. In addition, previous research has shown that members of each group tend to find a lot of similarities among themselves and are often surprised to realize they have much more in common than they would have initially thought.

Your results indicate that you belong to:

The Klees

In one sentence, tell us why you think you are a member of the Klees?

3. Pairing instructions (Klee) (Study 1)

The interaction section of this survey will now begin.
Please wait a moment while we pair you with another MTurk Worker.

You will be paired with the next MTurker who takes this survey and completes the demographics form and the Artistic Preference test.

Please note that you can either be paired with someone from the **same Artistic preference group** as you or someone from **a different Artistic preference group.**

As part of a research project in personality, we are interested to see how people with common or different characteristics work together online.

Your test determined you belong to the Klees. You can either be paired with another Klee, or with a Kandinsky.

The survey will advance automatically once someone has been found. Finding a match usually takes no longer than one or two minutes.

[22] For convenience, we will present only the materials for the Klee condition. For the Kandinsky condition, all materials were the same, with the exception that the word "Klee" was replaced by the word "Kandinsky".

4. Other (fictitious) person's profile[23] (Study 1)

We have successfully found another MTurk Worker!

You have been paired with:
A Klee

Please take a moment to get to know the individual you have been paired with through his/her demographic form.

Demographic information form:

1. Age: **32**
2. Gender: **${q://QID4/ChoiceGroup/SelectedChoices}**
3. Racial/ethnic category: **Caucasian/Non-Hispanic White**
4. Employment status: **Employed for wages**
5. Highest grade or year of school completed: **High school graduate**

[23] For convenience, we will present only the materials of the condition where a participant in the Klee condition interacted with a fictitious partner who was also a Klee. When the participant was in the Klee condition interacting with a (fictitious) Kandinsky partner, we replaced the expressions "the other Klee" for "the Kandinsky." When the participant was in the Kandinsky condition interacting with a Kandinsky partner, we used the expression "the other Kandinsky." When the participant was in the Kandinsky condition interacting with a Klee, we used the expression "the other Klee." All the remaining wording was the same.

5. Instructions (Study 1)

FIRST TASK DIRECTIONS

We will randomly assign one MTurker the role of answering a question first, while the other MTurker role is to give feedback to that answer. Then, the roles will switch.

When the MTurker who is answering submits his/her answer, it will become available to the MTurker who is giving feedback.

Then, it is the MTurker who is giving feedback's turn to provide feedback. After he/she submits feedback, it becomes available to the MTurker who is answering.

The MTurker who is answering can then decide to keep or change his/her initial answer, according to the feedback provided. Then, he/she submits a final answer, and the roles switch.

We ask you to limit your answers and feedback to two or three sentences.

6. Biased answer and participant's feedback (Study 1)

Question: **If you could change anything about the way you were raised, what would it be?**

Other Klee member's answer: **I was raised in a neighborhood with lots of blacks, so I was always scared. I wish I had been raised in a friendlier neighborhood with people who were more like me. Even today I don't like being around blacks, they are so aggressive all the time.**

Please give your feedback. Then, the other Klee member will have the opportunity to change his/her answer.
Your feedback will not be seen by anyone other than the other Klee member you are interacting with.

Your feedback:

7. Ratings of participants' own feedback (Study 1)

When you gave **feedback** to the other Klee member, to what extent did **you intend to**...

	Not at all					Very much
Be Supportive (of the other Klee member)	○	○	○	○	○	○
Be Protective (of the other Klee member)	○	○	○	○	○	○
Be Unpleasant	○	○	○	○	○	○
Be Informative	○	○	○	○	○	○
Be Considerate	○	○	○	○	○	○
Be Firm	○	○	○	○	○	○
Be Critical	○	○	○	○	○	○
Show Disagreement	○	○	○	○	○	○
Show Acceptance	○	○	○	○	○	○
Be Assertive	○	○	○	○	○	○
Show Displeasure	○	○	○	○	○	○
Be Helpful (to the other Klee member)	○	○	○	○	○	○
Be Strong	○	○	○	○	○	○

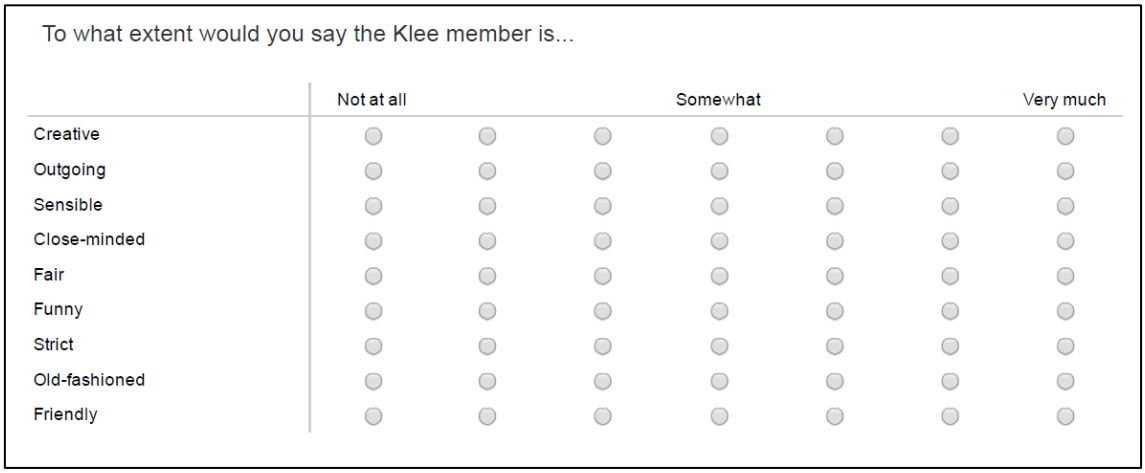
How much was **your feedback** intended to support the other Klee member because you agreed with what he/she said?

Not at all ○ ○ ○ Somewhat ○ ○ ○ Very much

How much was **your feedback** intended to confront the other Klee member because you disagreed with what he/she said?

Not at all ○ ○ ○ Somewhat ○ ○ ○ Very much

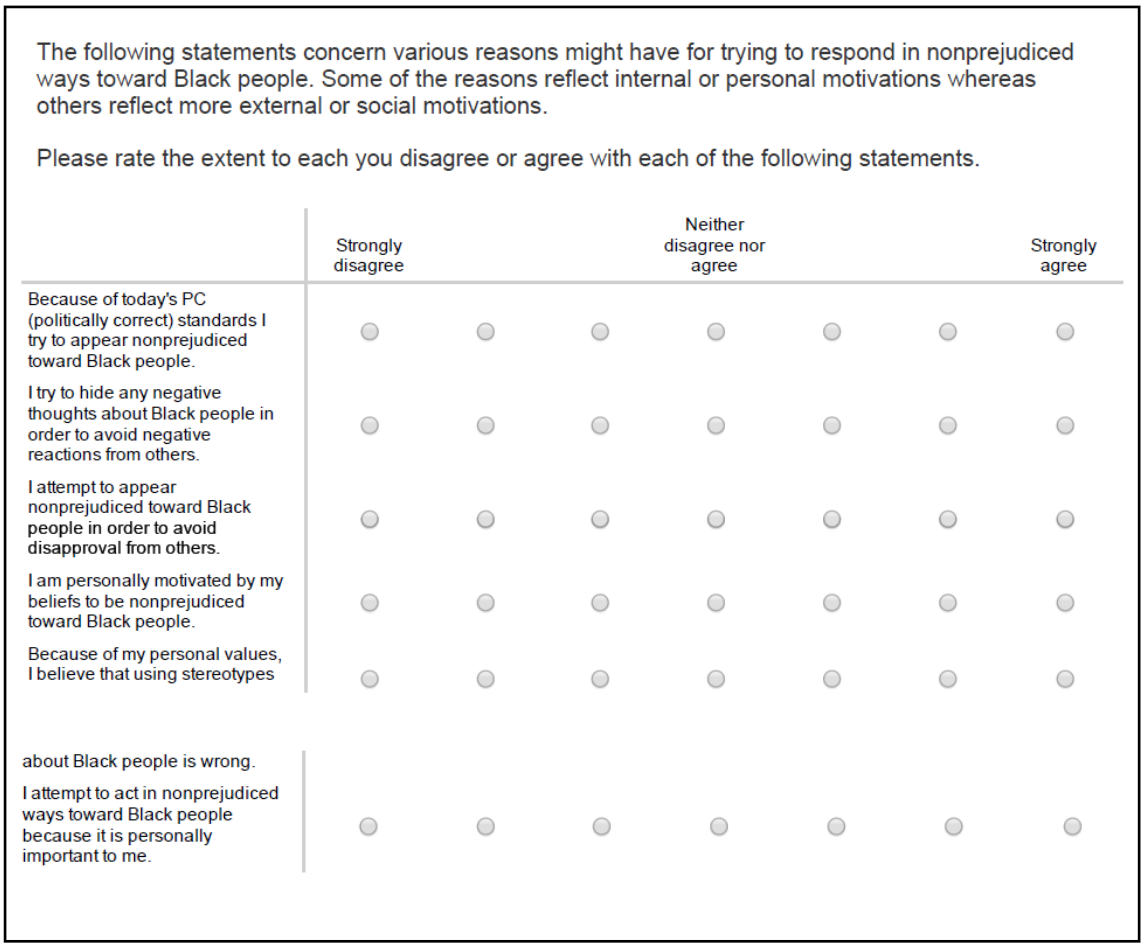
8. Attribution of positive, negative, and bias-related characteristics to the other (fictitious) person (Study 1)



To what extent would you say the Klee member is...

	Not at all			Somewhat			Very much
Creative	○	○	○	○	○	○	○
Outgoing	○	○	○	○	○	○	○
Sensible	○	○	○	○	○	○	○
Close-minded	○	○	○	○	○	○	○
Fair	○	○	○	○	○	○	○
Funny	○	○	○	○	○	○	○
Strict	○	○	○	○	○	○	○
Old-fashioned	○	○	○	○	○	○	○
Friendly	○	○	○	○	○	○	○

9. Internal and External motivation to respond without prejudice (Plant & Devine, 1998) (Study 1 and Study 2)



The following statements concern various reasons might have for trying to respond in nonprejudiced ways toward Black people. Some of the reasons reflect internal or personal motivations whereas others reflect more external or social motivations.

Please rate the extent to each you disagree or agree with each of the following statements.

	Strongly disagree			Neither disagree nor agree			Strongly agree
Because of today's PC (politically correct) standards I try to appear nonprejudiced toward Black people.	○	○	○	○	○	○	○
I try to hide any negative thoughts about Black people in order to avoid negative reactions from others.	○	○	○	○	○	○	○
I attempt to appear nonprejudiced toward Black people in order to avoid disapproval from others.	○	○	○	○	○	○	○
I am personally motivated by my beliefs to be nonprejudiced toward Black people.	○	○	○	○	○	○	○
Because of my personal values, I believe that using stereotypes about Black people is wrong.	○	○	○	○	○	○	○
I attempt to act in nonprejudiced ways toward Black people because it is personally important to me.	○	○	○	○	○	○	○

10. Instructions (Study 2)

Your next task is to help us characterize the members of the artistic perception groups.

In a previous study, we collected profiles of members of the two artistic perception groups, the Klees and the Kandinskys. Now, we ask participants like yourself to form an opinion of one of those profiles.

In the Artistic Preference Test, your results indicated you are a Klee. You can be assigned to form an opinion of a member of your own Artistic preference group (the Klees), or of a member of a different Artistic preference group (the Kandinskys).

Please note that your profile and opinions in this study are completely anonymous and will not be seen in the future by any member of the Klees or of the Kandinskys.

Click the "next" button to see the profile of the person of whom you will be forming an opinion.

11. Other (fictitious) person's profile, including the biased answer (Study 2)

Profile - Part I
Demographic information form:

Age: **32**
Gender: **${q://QID27/ChoiceGroup/SelectedChoices}**
Racial/ethnic category: **Caucasian/Non-Hispanic White**
Employment status: **Employed for wages**
Highest grade or year of school completed: **High school graduate**

Artistic preference group:
Klees

Profile - Part II
Open-ended questions:

Question 1: **When did you last sing to yourself? To someone else?**

Other Klee member's answer: **in the shower, like 3 days ago, I was so happy that day! I can't remember the last time I sang to some1 else, maybe like a year ago. I actually really like to sing, I should sing to other people more often**

Question 2: **Is there something that you've dreamed of doing for a long time?**

Other Klee member's answer: **I always wanted to travel around the globe for a year. Just put the minimal amount of things on a backpack and go, without much plans or a set destination. See where the world takes me.**

Question 3: **Given the choice of anyone, whom would you want as a dinner guest?**

Other Klee member's answer: **I'm not a very good cook, so I'd feel bad inviting anyone over for dinner.**

Question 4: **If you could change anything about the way you were raised, what would that be?**

Other Klee member's answer: **I was raised in a neighborhood with lots of blacks, so I was always scared. I wish I had been raised in a friendlier neighborhood with people who were more like me. Even today I don't like being around blacks, they are so aggressive all the time.**

Question 5: **Complete this sentence: "I wish I had someone with whom I could share ... "**

Other Klee member's answer: **my unconditional caring.**

12. Attribution of positive, negative, and bias-related characteristics to the other (fictitious) person (Study 2)

To what extent would you say the other Klee member is...

	Not at all			Somewhat			Very much
Old-fashioned	○	○	○	○	○	○	○
Funny	○	○	○	○	○	○	○
Insecure	○	○	○	○	○	○	○
Ignorant	○	○	○	○	○	○	○
Friendly	○	○	○	○	○	○	○
Sensitive	○	○	○	○	○	○	○
Outgoing	○	○	○	○	○	○	○
Creative	○	○	○	○	○	○	○
Naive	○	○	○	○	○	○	○
Fair	○	○	○	○	○	○	○
Strict	○	○	○	○	○	○	○
Close-minded	○	○	○	○	○	○	○
Fearful of change	○	○	○	○	○	○	○

Appendix C

Materials used in Study 1, Study 2 and Study 3 of Chapter IV

1. Instructions of the self-presentation paradigm (Study 1, Study 2 and Study 3).

Social approval condition:

Vamos apresentar-lhe algumas situações sociais. A sua tarefa é responder a cada situação de modo a dar uma imagem **POSITIVA** de si, isto é, de forma a que **quem ler as suas respostas as aprove e fique a gostar de si.**

Social disapproval condition:

Vamos apresentar-lhe algumas situações sociais. A sua tarefa é responder a cada situação de modo a dar uma imagem **NEGATIVA** de si, isto é, de forma a que **quem ler as suas respostas as desaprove e fique a não gostar de si.**

Standard condition:

Vamos apresentar-lhe algumas situações sociais. A sua tarefa é responder a cada situação de acordo com **A SUA OPINIÃO**. Não existem respostas certas nem erradas pois todas representam a sua **opinião honesta sobre cada uma das situações.**

2. Scenarios used in the self-presentation paradigm (Study 1)

Está no metro e ouve um português fazer um comentário preconceituoso contra imigrantes.

Ignorava o comentário ou dizia à pessoa portuguesa que o comentário foi preconceituoso?

Ignorava o comentário						Dizia que foi preconceituoso
1	2	3	4	5	6	7

Está num jardim público e ouve um português a fazer um comentário preconceituoso contra imigrantes.

Ignorava o comentário ou dizia à pessoa portuguesa que o comentário foi preconceituoso?

Ignorava o comentário						Dizia que foi preconceituoso
1	2	3	4	5	6	7

Está numa fila de supermercado e ouve um português fazer um comentário preconceituoso contra imigrantes.

Ignorava o comentário ou dizia à pessoa portuguesa que o comentário foi preconceituoso?

Ignorava o comentário						Dizia que foi preconceituoso
1	2	3	4	5	6	7

Está num café com amigos e um deles, que é português, faz um comentário preconceituoso contra imigrantes.

Ignorava o comentário ou dizia ao seu amigo que o comentário foi preconceituoso?

Ignorava o comentário						Dizia que foi preconceituoso
1	2	3	4	5	6	7

Está num almoço de família e um dos seus familiares faz um comentário preconceituoso.

Ignorava o comentário ou dizia ao seu familiar que o comentário foi preconceituoso?

Ignorava o comentário						Dizia que foi preconceituoso
1	2	3	4	5	6	7

Está a fazer um trabalho de grupo e um dos seus colegas, que é português, faz um comentário preconceituoso contra imigrantes.

Ignorava o comentário ou dizia ao seu colega que o comentário foi preconceituoso?

Ignorava o comentário						Dizia que foi preconceituoso
1	2	3	4	5	6	7

3. Attention check (Study 1 and Study 2)

Queremos saber se se lembra das instruções que lhe demos na primeira parte deste questionário. Por favor responda à questão seguinte sem voltar a ver as páginas anteriores, mesmo que não se lembre exatamente das instruções.

Na primeira parte do questionário, pedimos-lhe que:

a) Respondesse de acordo com a sua opinião.
b) Respondesse de forma a dar uma imagem positiva de si.
c) Respondesse de forma a dar uma imagem negativa de si.

4. Scenarios used in the self-presentation paradigm (Study 2 and Study 3). Stranger condition is in brackets.

Está numa esplanada de um café com os seus amigos. Estão todos a conversar quando passa um imigrante negro na rua. Um dos seus amigos [Um desconhecido na mesa do lado], que é português, diz: "Não percebo como é que continuam a deixar vir para cá imigrantes que só vêm estragar o que é nosso."

Quão provável seria que fizesse cada uma das seguintes ações …

	Nada provável					Muito provável
Dizer ao seu amigo [ao desconhecido] que o comentário foi preconceituoso.	1	2	3	4	5	6
Ignorar o comentário do seu amigo [do desconhecido].	1	2	3	4	5	6
Mostrar ao seu amigo [ao desconhecido] que o comentário foi incorreto.	1	2	3	4	5	6
Mudar de assunto.	1	2	3	4	5	6

5. Threat to the image of a the friends' group and threat to the image of the Portuguese (Study 3)

Nesta situação, em que medida acha que o seu amigo [o desconhecido]…

	Discorda muito					Concorda muito
Ameaçou a imagem do seu grupo de amigos.	1	2	3	4	5	6
Ameaçou a reputação do seu grupo de amigos.	1	2	3	4	5	6
Ameaçou a reputação dos portugueses.	1	2	3	4	5	6
Ameaçou a imagem dos portugueses.	1	2	3	4	5	6

6. Costs of confronting (Study 3)

Acha que seria fácil ou difícil dizer ao seu amigo [ao desconhecido] que o comentário foi preconceituoso?

Muito fácil					Muito difícil
1	2	3	4	5	6

Acha que seria seguro ou arriscado dizer ao seu amigo [ao desconhecido] que o comentário foi preconceituoso?

Muito seguro					Muito arriscado
1	2	3	4	5	6

Acha que teria consequências positivas ou negativas dizer ao seu amigo [ao desconhecido] que o comentário foi preconceituoso?

Muitas consequências positivas					Muitas consequências negativas
1	2	3	4	5	6